



**UNIVERSIDAD NACIONAL AUTÓNOMA  
DE MÉXICO**

---

---

**FACULTAD DE CIENCIAS**

**Diversidad microbiónal asociada a datos de RADSeq  
de escarabajos descortezadores del género *Dendroctonus***

**T E S I S**

**QUE PARA OBTENER EL TÍTULO DE:**

**BIÓLOGO**

**P R E S E N T A:**

**DANIEL ENRIQUE ACOSTA GARCÍA**

**DIRECTOR DE TESIS:  
M. EN C. NATALIA DELGADO MACHUCA  
2022**

Ciudad de México, 2022



Universidad Nacional  
Autónoma de México

Dirección General de Bibliotecas de la UNAM

**Biblioteca Central**



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.



1. Datos del alumno

Acosta García Daniel Enrique  
Universidad Nacional Autónoma de México  
Facultad de Ciencias  
Biología  
308532740

2. Datos del tutor

M. en C. Natalia Delgado Machuca  
Universidad Nacional Autónoma de México  
Instituto de Biología  
Departamento de Zoología

3. Datos del sinodal 1

Dr. en Q.B Francisco Armendariz Toledano  
Universidad Nacional Autónoma de México  
Instituto de Biología  
Departamento de Zoología

4. Datos del sinodal 2

Dr. Juan Alfredo Hernández García  
Instituto Politécnico Nacional  
Escuela Nacional de Ciencias Biológicas  
Departamento de Microbiología

5. Datos del sinodal 3

PhD Alicia Mastretta Yanes  
Consejo Nacional de Ciencia y Tecnología  
Comisión Nacional para el Uso y Conocimiento de la Biodiversidad  
Coordinación de Agrobiodiversidad y Recursos Biológicos

6. Datos del sinodal 4

Dr. Luis David Alcaraz Peraza  
Universidad Nacional Autónoma de México  
Facultad de Ciencias  
Departamento de Biología Celular

Datos del trabajo escrito

Diversidad microbiana asociada a datos de RADSeq de escarabajos descortezadores del género *Dendroctonus*  
Número de Páginas 108  
Año 2022

## AGRADECIMIENTOS

Le agradezco a mi familia, a mis padres por siempre apoyarme y permitirme perseguir mis sueños con plena confianza. A mi hermano por darme ánimos y por escucharme en mis ratos de más estrés y dudas.

A mis abuelos que siempre me motivaron, a mis tías, Hilda, Norma y a mi prima Cecile, sin cuyos abrazos y cariño no habría llegado tan lejos.

A mi perrita Frida por su cariño y lealtad incondicional, su cercanía me hace sentirme muy agradecido por nuestra condición simbiótica.

A mis amigos Diego, Sofía, Emilio, Cybelle y Eduardo. Les quiero como a nadie, no concibo mi juventud sin nuestros días de gloria y nuestras metidas de pata.

A mis amigos y compañeros de carrera, Uriel, Shamady, Montse, Memo y Ari. Todas esas conversaciones en la facultad y experiencias en campo las llevo conmigo, sin ustedes no hubiera disfrutado tanto perderme en este camino científico a lo largo de estos años.

A Marisol Navarro, por guiarme a lo largo de la carrera y continuar siendo una inspiración en la bioinformática y en la ciencia. Te quiero y te agradezco por abrir camino.

A mi asesora Natalia, a la Dra. Rubí Meza y a todos los profesores del Taller “Herramientas genómicas para el estudio de la biodiversidad y evolución”. Gracias por confiarme en este proyecto y brindarme las enseñanzas y las direcciones necesarias.

A la Coordinación de Supercómputo de la Dirección General de Tecnologías de la Información y Comunicación de la UNAM por habernos dado la oportunidad de trabajar en su clúster de supercómputo, Miztli. También a Irving y a la Dra. Alicia Mastretta de CONABIO, por darnos acceso a su clúster cuando este proyecto estaba iniciando. Sin su apoyo esta tesis no se habría llevado a cabo.

A la Dra. Camille Truong, por haberme introducido al mundo de la bioinformática, de la ecología microbiana y de la simbiosis con insectos.

# ÍNDICE

I. INTRODUCCIÓN	8
A. Escarabajos del género <i>Dendroctonus</i>	8
B. Comunidades microbianas asociadas a <i>Dendroctonus</i> spp.	14
1. Bacterias	14
2. Fungi: Hongos miceliales y levaduras	17
3. Nemátodos	20
C. Métodos de estudio de la microbiota	20
D. Métodos de secuenciación de representación reducida del genoma	23
II. JUSTIFICACIÓN	24
III. OBJETIVOS	25
IV. MÉTODOS	26
A. Obtención de datos	26
B. Ensamble de datos	27
C. Determinación taxonómica y análisis	31
V. RESULTADOS	34
A. Ensamble de datos	34
B. Determinación taxonómica	39
C. Taxa de interés	47
D. Tamaño de los loci	53
E. Distribución de loci entre las diferentes especies de <i>Dendroctonus</i>	57
VI. DISCUSIÓN	67
A. Consideraciones metodológicas	67
1. Modelo biológico y diseño experimental	67
2. Ensamble y recuperación de loci	69
3. Determinación taxonómica	71
4. Taxa <i>raros</i>	74
5. Comparaciones de diversidad	75
B. Consideraciones ecológicas	76
1. Fungi	76
2. Plantae	78
3. Bacteria	80
4. Nematoda	84
C. Perspectivas	85
1. Recomendaciones del muestreo y la obtención de secuencias	85
2. Recomendaciones técnicas	87
VII. CONCLUSIONES	88
VIII. REFERENCIAS	90



# INTRODUCCIÓN

## Escarabajos descortezadores del género *Dendroctonus*

El término "escarabajo descortezador" se utiliza de manera coloquial para referirse a las especies de la subfamilia Scolytinae, familia Curculionidae. Estos escarabajos se caracterizan por completar parte de su ciclo de vida principalmente dentro de un árbol deteriorando a su hospedero al alimentarse de diferentes tejidos vegetales como la corteza, el xylema, el floema, las semillas y la miel, e incluso de estructuras fúngicas como el micelio (Knizek y Beaver, 2007). Esta capacidad de desarrollarse dentro de los tejidos vegetales se ha asociado frecuentemente con la presencia de hongos mutualistas, ya que ayudan a degradar compuestos complejos del árbol de los que se alimentan los coleópteros. Tal es el caso de los escarabajos del género *Dendroctonus* Erichson, 1836, los cuales se alimentan principalmente de floema de pináceas de los géneros *Larix*, *Picea*, *Pinus* y *Pseudotsuga* por medio de diferentes tipos de comportamiento: agresivas y saprobias tempranas (Six *et al.*, 2004). Los miembros del género *Dendroctonus*, se encuentran entre las especies más importantes de escarabajos descortezadores desde el punto de vista económico y ecológico, ya que afectan fuertemente a los ecosistemas forestales de América del Norte y Central (Bentz, 2008).

A pesar de que la mortalidad de los árboles como resultado de los brotes de *Dendroctonus* ha afectado fuertemente a la industria maderera, los eventos de perturbación causados por las especies nativas de escarabajos de la corteza son también importantes impulsores de la sucesión forestal y la salud general de estos ecosistemas (Samman y Logan, 2000). Los integrantes del género *Dendroctonus* destacan entre los parásitos descortezadores más agresivos de los bosques de pino de Norte y Centroamérica (Wood, 1982). Su relevancia taxonómica disminuye en Europa y en China debido a la presencia de únicamente dos especies nativas y una introducida (Grégoire, 1988, Wang *et al.*, 2010). El impacto ecosistémico y económico que tienen los hábitos agresivos de algunas de estas especies han dado pie a su estudio desde diversas perspectivas, principalmente en la zona templada de Norteamérica (Six y Bracewell, 2015).

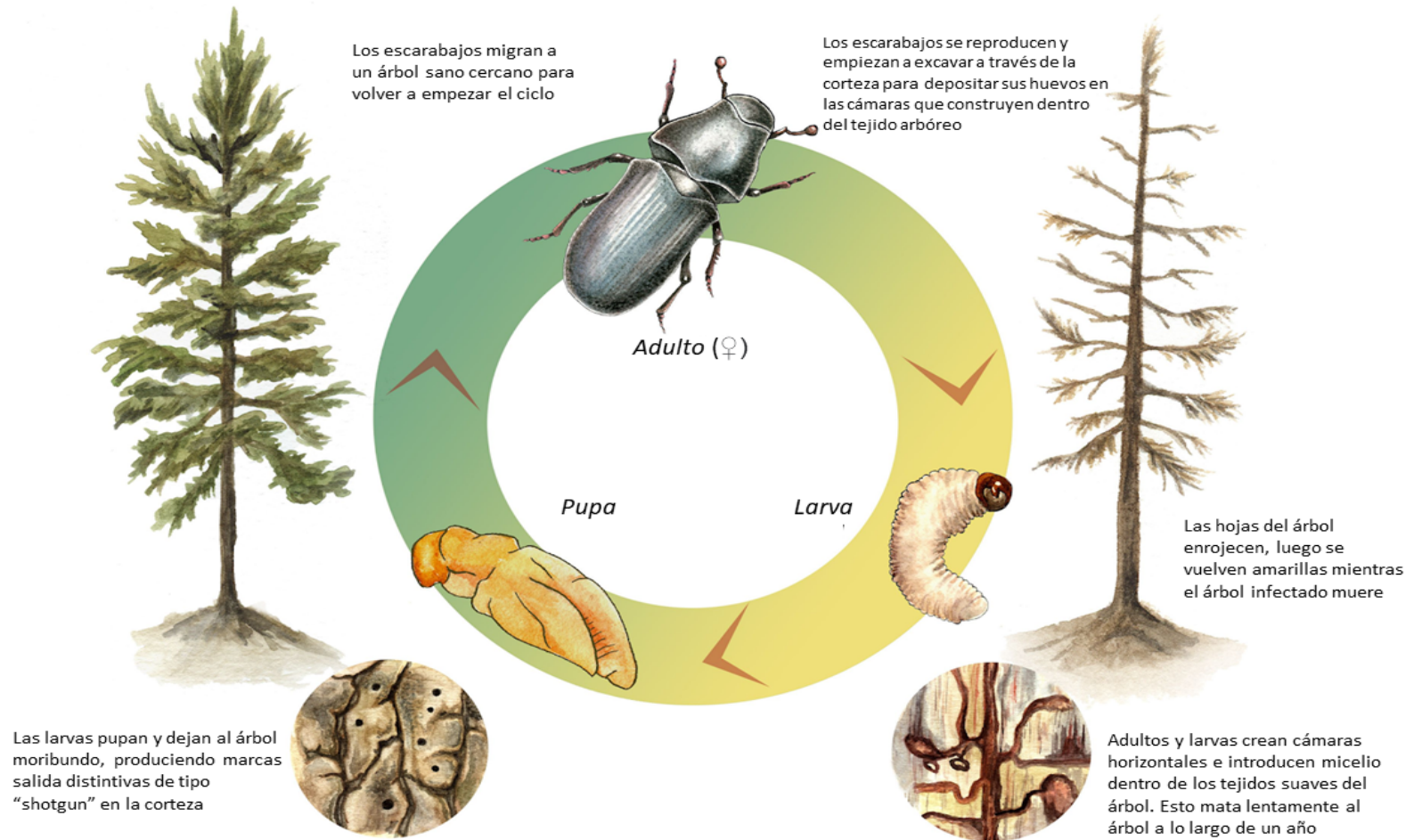
El género *Dendroctonus* está compuesto por veintiún especies, incluyendo a la especie críptica *D. barberi* Hopkins la cual ha sido recientemente delimitada de *D. brevicomis* (Valerio-Mendoza *et al.*, 2019; Sullivan *et al.*, 2021) Históricamente las especies de este género se han clasificado en cinco o seis grupos con base en sus características morfológicas (Wood, 1963; Reeve *et al.*, 2012). Estos grupos, en su mayor parte, se han conservado como clados o subclados con base en estudios filogenéticos (Victor y Zúñiga, 2016). El género en su totalidad destaca entre el resto de la



subfamilia Scolytinae debido a la presencia de un proceso epistomal en la cabeza, es decir, una maza antenal aplanada formada por cuatro antenómeros (segmentos que forman a la antena) y las proyecciones flagelares de cinco antenómeros. El tamaño promedio de los integrantes del género *Dendroctonus* ronda entre 2.4 mm a 4.0 mm dentro del complejo de especies más pequeñas, y hasta entre 7 y 7.3 mm en el complejo de especies más grandes; sin embargo estos intervalos pueden variar en gran medida entre poblaciones y localidades debido a la disponibilidad de alimento y a otras variables ambientales (Armendáriz-Toledano *et al.*, 2018). El color de los ejemplares adultos se emplea para distinguir su grado de desarrollo, el cual también puede variar geográfica e interpoblacionalmente. En todo el género, los escarabajos más jóvenes tienden a presentar coloraciones más claras antes de alcanzar la madurez, mientras que algunas especies se caracterizan por presentar una coloración homogénea en la adultez (*D. approximatus*, *D. brevicomis* y *D. parallelcollis*), y en otras, la cabeza y el pronoto tórax anterior son más oscuros que el abdomen y los élitros (*e. g.*, *D. frontalis*, *D. jeffreyi*, *D. mesoamericanus*, *D. mexicanus*, *D. ponderosa*, *D. pseudotsugae*, *D. rhizophagus*, *D. valens* y *D. vitei*) (Armendáriz-Toledano *et al.*, 2018).

El ciclo de vida general del género *Dendroctonus*, esquematizado con la especie *D. ponderosae* en la Figura 1, se divide en la etapa de Colonización (1) de especies pináceas, a menudo iniciada por las hembras, Reproducción y desarrollo (2) dentro del sistema de galerías que construyen en el floema arbóreo y Emergencia de la nueva generación de adultos (3). A través de ataques subsiguientes de más individuos de la misma especie, estos logran entrar al floema del árbol hasta alcanzar los tejidos donde construyen cámaras de cría para depositar sus huevos. Los adultos y las larvas de algunas especies construyen cámaras nupciales, túneles parentales y túneles larvales. En las especies que hay túneles parentales, algunas son horizontales, como *D. ponderosae*, *D. parallelcollis*, *D. pseudotsugae*, y en otras suele ser perpendicular. Los sistemas de galerías más complejos a menudo conservan patrones arquitectónicos característicos de cada especie (Armendáriz-Toledano *et al.*, 2018). En el sistema de galerías se puede encontrar una gran variedad de hongos transportados por los escarabajos, principalmente ascomicetos (Phylum Ascomycota), los cuales debido a su capacidad para degradar la madera se asocian con la patogénesis y la muerte gradual del árbol (Harrington, 19993; Six, 2003). Cuando las crías superan la etapa de pupa, a menudo después de períodos de metamorfosis de hasta más de un año, abandonan el árbol infectado produciendo hoyos de salida distintivos en la corteza del árbol moribundo y posteriormente migrando para infectar un nuevo árbol (Twiddy, 2010).

## Ciclo de vida del escarabajo de pino de montaña, *D. ponderosae*



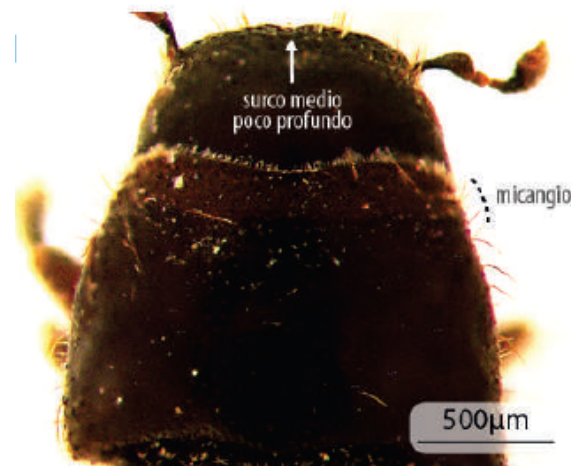
J.A. Leatherman, I. Aguayo, and T.M. Mehall

Sean Twiddy 2010

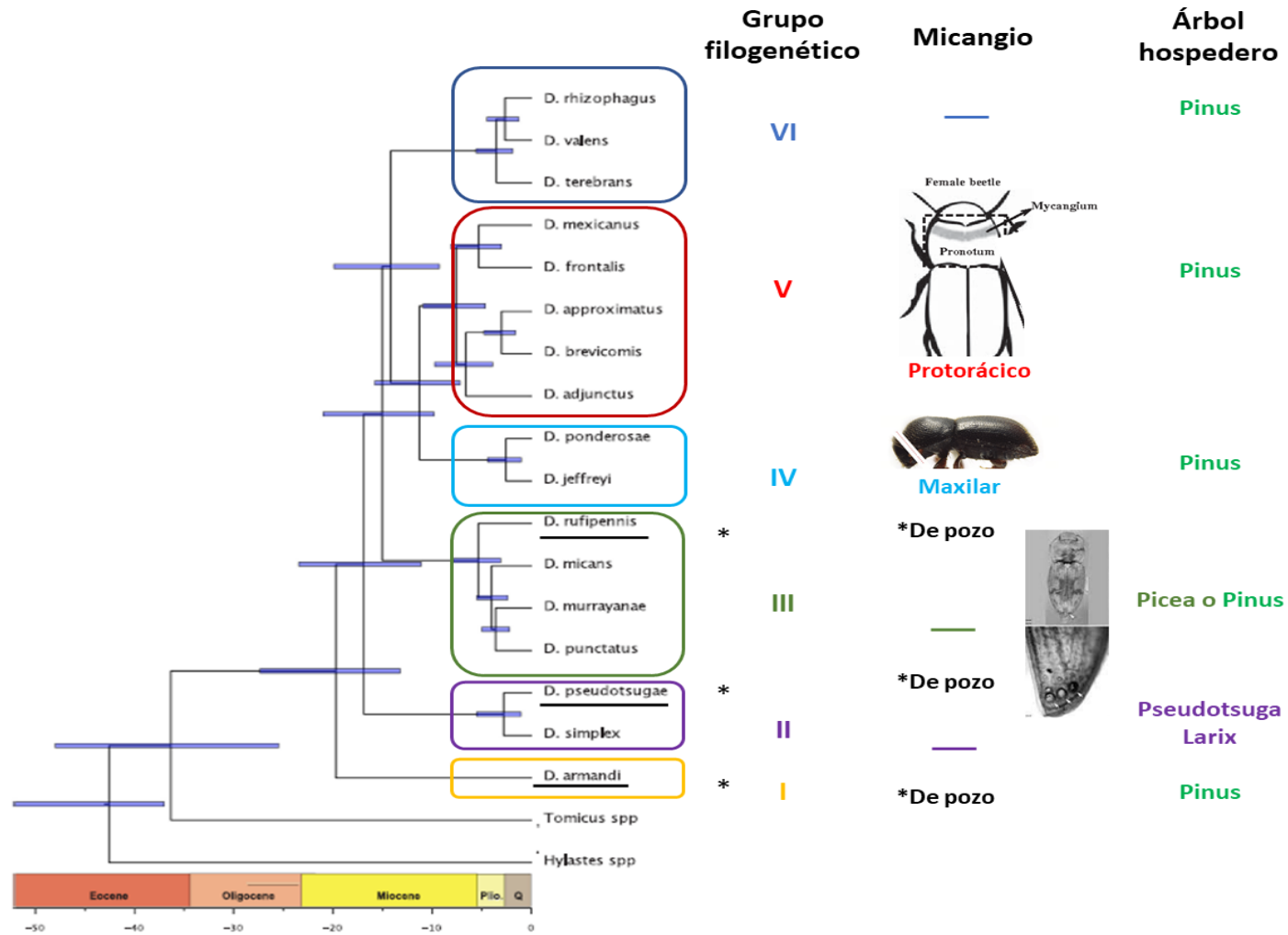
**Figura 1.** Ciclo de vida de *Dendroctonus ponderosae*. Al igual que otras especies del género *D. ponderosae* se caracteriza por colonizar árboles durante la etapa adulta (1) y

construir sistemas de galerías en el floema de los árboles con el fin de criar sus larvas en ellas (2). Cuando las pupas concluyen la metamorfosis, los escarabajos adultos abandonan las galerías para reiniciar el ciclo y dejan a su paso hoyos en la corteza que el árbol suele tapar con resina (3). Tomada de Twiddy, 2010.

Prácticamente todos los escarabajos descortezadores portan esporas de múltiples hongos en su cuerpo, durante algún momento de su ciclo de vida (Six y Klezpig, 2004). Además, algunas especies de este género presentan una simbiosis obligada, definida por el desarrollo de una estructura glandular llamada micangio, la cual secreta sustancias grasosas y cerosas que permiten la proliferación y el almacenamiento de hongos filamentosos simbiotes (Six y Klezpig, 2004; Six, 2019). El micangio ha sido empleado como un carácter filogenético dentro de este género, ya que puede presentarse de diferentes maneras: como un micangio de pozo localizado en los élitros o en el pronoto, como un par de sacos localizados cerca de las mandíbulas (micangio maxilar), o bien como un callo transversal elevado en la región anterior de las hembras (micangio protorácico) (Fig 2) (Armendáriz-Toledano *et al.*, 2018; Six y Klezpig, 2004).



**Figura 2.** Vista dorsal del micangio de la hembra de *D. adjunctus*. El micangio se observa como una línea blanca en medio de la cabeza y el tórax debido al micelio que se almacena dentro de la cavidad glandular. Tomada de Armendáriz-Toledano *et al.*, 2018



**Figura 3.** Filogenia de diecisiete especies del género *Dendroctonus* a partir de datos de RADSeq. Las especies con micangio protorácico pertenecen al grupo filogenético V y se encuentran resaltadas de color rojo, próximas al grupo IV resaltado de color amarillo e integrado por las dos especies con micangio maxilar. Las especies en las que se han

encontrado “micangios de pozo” se resaltaron con una línea negra. Además al costado de cada grupo se ha indicado el género de Pinaceae que infectan. Se incluyeron a dos especies de los géneros *Tomicus* y *Hylastes* como grupos externos. (Modificada de Godefroid *et al.*, 2019)

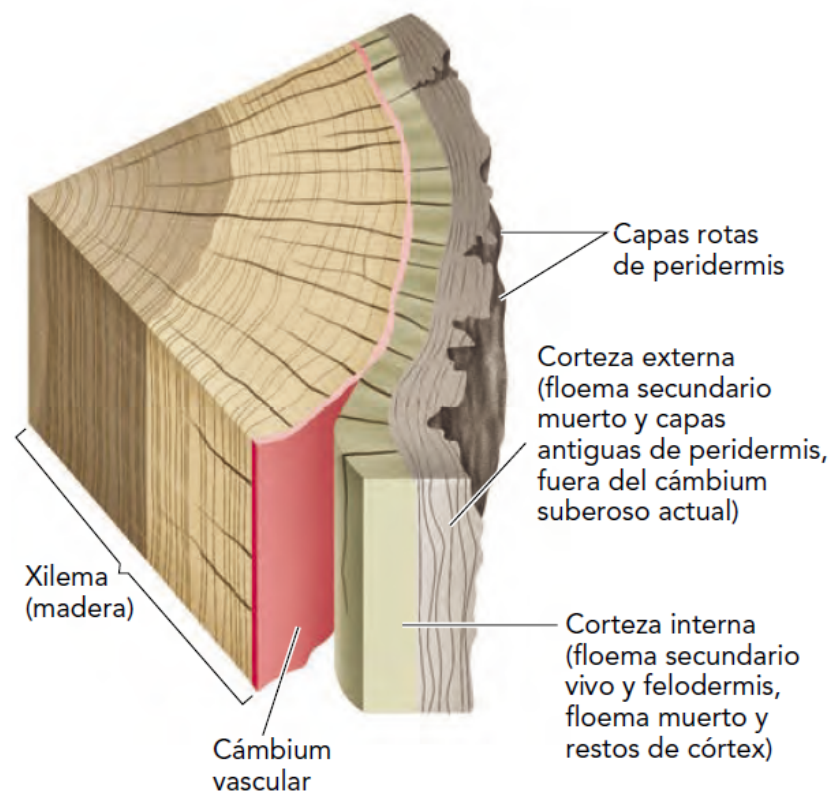
Los grupos filogenéticos representados en el trabajo de Godefroid y colaboradores de 2019 ya habían sido reconocidos a partir de algunos caracteres como el cariotipo, el número de cromosomas, el sistema de determinación sexual, el comportamiento larval y el género de árbol de la familia Pinaceae que infecta. Este último atributo divergió en el grupo 2 (*D. simplex* y *D. pseudotsugae*), el cual coloniza a *Larix* y a *Pseudotsuga*, y en el grupo 3 (*D. micans*, *D. murrayanae*, *D. rufipennis* y *D. punctatus*), que infecta al género *Picea*; mientras que el resto de los grupos infectan a *Pinus* (Six y Klepzig, 2004; Zúñiga *et al.*, 2002).

El desarrollo de los diferentes tipos de micangio no refleja homologías evolutivas entre las especies del género ya que aparecieron en múltiples ocasiones en los linajes pertenecientes a la filogenia del género (Victor y Zúñiga, 2016; Six, 2019). Además aunque el concepto de *micangio* se ha empleado de manera general para referirse a cualquier estructura empleada para el transporte y almacenamiento temporal de hongos simbiotes, como el micangio de pozo, se le denomina micangios verdaderos a las estructuras saculares con una especialización morfológica estricta y documentada para el transporte y almacenamiento de hongos simbióticos (Victor y Zúñiga, 2016). El micangio de pozo ha sido descrito únicamente en algunas especies de los primeros tres grupos (*D. armandi*, *D. pseudotsugae* y *D. rufipennis*), y aunque ha sido menos estudiado que los otros tipos de micangio, se ha confirmado que su localización le permite atrapar y proteger las esporas oportunamente, antes de la salida del escarabajo adulto de su árbol natal (Six y Klepzig, 2003; Six, 2019; Godefroid *et al.*, 2019).

Los micangios verdaderos se encuentran en dos grupos resaltados en amarillo y rojo en la Fig. 3, el primero contiene a las dos especies con micangio maxilar presente en ambos sexos *D. ponderosae* y *D. jeffreyi*, y el segundo contiene a aquellas especies cuyas hembras desarrollan un micangio protorácico: *D. frontalis*, *D. mexicanus*, *D. adjunctus*, *D. approximatus*, *D. brevicomis*, *D. vitei* y *D. mesoamericanus*. Estas últimas dos especies no fueron incluidas en la filogenia de la figura 3 (Six y Klepzig, 2004, Armendáriz-Toledano *et al.*, 2018; Six, 2019; Godefroid *et al.*, 2019).

El desarrollo de mayoría de las especies de *Dendroctonus* depende de un tejido vegetal interno de menor complejidad estructural conocido como floema, sin embargo, la importancia de los hongos simbiotes en la nutrición de las larvas y pupas aumenta en las especies que migran a la

corteza externa durante su metamorfosis, agrupadas filogenéticamente en el grupo V y caracterizadas por la presencia de un micangio protorácico. En este clado, los simbiontes han sido reconocidos como co-constructores de nicho debido a su capacidad de transportar nutrientes desde el floema y de degradar los compuestos vegetales de estructura compleja contenidos en la corteza externa como la lignina, la celulosa y las hemicelulosas (Six, 2019, Six and Paine, 1996). En algunas especies con micangio protorácico, que no presentan una independencia completa del floema, se cree que los hongos micangiales fungen como una fuente de vitaminas, esteroides y otros nutrientes que, potencialmente influyen en el desarrollo de la larva y la reducción de la competencia contra otros insectos mediante la construcción de galerías más cortas (Six, 2012; Farrel *et al.*, 2001).



**Figura 4.** Corte transversal de la corteza externa e interna del eje vascular.. El crecimiento secundario o diametral con este tipo de arreglo tisular se encuentra en angiospermas dicotiledóneas y gimnospermas, como los pinos colonizados por *Dendroctonus*. Modificada de Nabors, M., 2006.

Además de los taxa de hongos asociados frecuentemente a *Dendroctonus*, se ha propuesto que las comunidades bacterianas presentes en las diferentes etapas del ciclo de vida de este género pueden tener un papel crucial en la construcción de su nicho (Briones-Roblero, 2017; Hernández - García, 2017; Adams, 2009). Otros simbiontes que han sido encontrados y que se han propuesto como componentes adicionales de la interacción árbol - hongo - escarabajo incluyen ácaros y nemátodos,

cuya relación ecológica como mutualistas, parásitos o comensales no ha sido estudiada en la mayor parte del género (Klepzig, 2003; Kinn, 1981).

El estudio de los simbioses asociados con *Dendroctonus* ha cobrado gran relevancia durante las últimas décadas en zonas templadas de América del Norte (Lee *et al.*, 2006; Roe *et al.*, 2011; Skelton *et al.*, 2018), y solo hay pocos estudios centrados en áreas tropicales. Además, los bosques templados han experimentado modificaciones importantes debido al cambio climático y la mala gestión de los recursos, lo que ha resultado en infestaciones importantes de escarabajos, conduciendo a la reducción de las poblaciones de pinos y a la afectación de la industria maderera (Loeffer y Anderson, 2018; La Tribuna, 2019). Históricamente, en México se produjeron grandes eventos de brotes de escarabajos en la parte oriental del Eje Neovolcánico Transversal, sumados a eventos recientes en elevaciones más bajas y durante el verano (del-Val y Sáenz-Romero, 2017). En este contexto, el estudio de las asociaciones con este escarabajo, es clave para desarrollar nuevas estrategias para el manejo de las especies de escarabajos descortezadores a lo largo de Norteamérica.

## **Comunidades microbianas asociadas a *Dendroctonus* spp.**

### **Bacterias**

La visión clásica de simbiosis, definida por el botánico y micólogo Albert Bernhard Frank como una relación íntima entre dos organismos de diferentes especies (Gontier, 2016), ha sido complementada durante las últimas décadas por los conceptos emergentes del estudio de la *microbiota*, entendida como la comunidad de microorganismos asociada específicamente a un hospedero (Aschenbrenner *et al.*, 2016).

El impacto de los rebrotes y la distribución geográfica del género *Dendroctonus* varía dependiendo del huésped descortezador y el estado del ecosistema que habita. Esta diferencia en la importancia ecológica y socio-económica, ha implicado que algunas especies están mejor estudiadas y documentadas que otras, incluyendo los esfuerzos por conocer y caracterizar al conjunto de los genes asociados a la comunidad de bacterias simbioses (conocido como microbioma bacteriano o bacterioma) compartido entre múltiples especies del género. El trabajo más destacado en este tema fue publicado por Hernández-García (2017), en el cual se investigó el bacterioma intestinal de 13 especies mexicanas (*D. adjunctus*, *D. approximatus*, *D. brevicomis*, *D. frontalis*, *D. jeffreyi*, *D.*

*mesoamericanus*, *D. mexicanus*, *D. parallelocollis*, *D. ponderosae*, *D. pseudotsugae*, *D. rhizophagus*, *D. valens* y *D. vitei*).

Los estudios más extensivos para la caracterización de la microbiota surgieron a partir de las técnicas de secuenciación masiva en paralelo (SMP) durante las últimas décadas (White *et al.*, 2016). A partir de ellos se ha acordado el uso de Unidades Taxonómicas Operacionales (OTUs) para determinar especies u otros niveles taxonómicos mediante un porcentaje de similitud determinado, regularmente de ~97%, entre las secuencias de la microbiota y una base de datos de referencia (Brietawaser *et al.*, 2019).

Hernández-García reporta 5 phyla bacterianos presentes en todas las especies de *Dendroctonus*, destacando Proteobacteria como el más abundante, seguido de Firmicutes, Fusobacteria, Actinobacteria y Deinococcus-Thermus. El microbioma central estricto, presente en el 100% de las diferentes especies de *Dendroctonus* y sus réplicas, estuvo constituido por seis géneros (*Enterobacter*, *Pantoea*, *Pseudomonas*, *Rahnella*, *Raoultella* y *Serratia*), mientras que el microbioma central relajado (60%) incluyó ocho taxones adicionales *Acinetobacter*, *Providencia*, *Stenotrophomonas*, *Erwinia*, *Kluyvera*, *Paenibacillus*, *Ralstonia* y algunos miembros no clasificados de la familia Enterobacteriaceae (Hernández-García *et al.*, 2017). Las especies de *Dendroctonus* que presentaron mayor abundancia de “especies bacterianas”, fueron *D. brevicomis*, *D. jeffreyi* y *D. valens* con 99, 98 y 87.5 OTU’s observadas respectivamente, mientras que las especies con una menor abundancia de OTU’s fueron *D. approximatus* y *D. pseudotsugae* con 30 y 25 OTU’s respectivamente. Además en este estudio se realizaron comparaciones de la diversidad bacteriana encontrada en las diferentes especies de *Dendroctonus* (diversidad  $\beta$ ), encontrando diferencias significativas entre la riqueza de especies esperada, la dominancia de géneros bacterianos y la diversidad filogenética de las comunidades de bacterias (Hernández-García *et al.*, 2017).

Los resultados publicados en un artículo de Hernández-García del 2018, han confirmado la presencia de la mayor parte de los OTUs bacterianos identificados como componentes del microbioma central de *Dendroctonus* en muestras de *D. valens* y *D. mexicanus*, las cuales se conservan en su mayor parte en poblaciones independientes distribuidas a lo largo de los sistemas montañosos de México. La investigación demostró que en estas especies, las comunidades bacterianas de distintas poblaciones aisladas geográficamente varían únicamente por la presencia de ciertos taxa bacterianos poco abundantes o “raros”. No fue posible identificar una correlación entre el nivel de similitud de las comunidades bacterianas calculado a partir de su posición en árboles filogenéticos



construidos con las OTUs encontradas (distancia filogenética o índice UniFrac), con la distancia geográfica que las separaba (Hernández-García *et al.*, 2018).

Otro de los enfoques disponibles para la caracterización de las comunidades microbianas asociadas a insectos, consiste en comparar el bacterioma de una región externa o superficial como la cutícula (denominado ectomicrobioma) con el de una región aislable, como una estructura de almacenamiento de simbiontes, o bien con el bacterioma de la región interna, comúnmente limitada por el sistema digestivo debido a la gran diversidad microbiana que este alberga y a la relevancia fisiológica que suelen tener estas comunidades en el hospedero. A través del estudio de las comunidades bacterianas de *D. simplex*, se encontraron diferencias importantes entre los ambientes incluidos: partiendo de 210 OTUs (95 % de similitud) registrados como componentes de la endomicrobiota, 15 integrantes de las comunidades asociadas a la cutícula y 73 a partir de las galerías que construyen estos escarabajos en la corteza de diferentes especies de pino (Durand *et al.*, 2015). La mayor parte de los OTUs encontrados pertenecen a los phyla Proteobacteria, Acidobacteria, Bacteroidetes y Firmicutes. Dentro de la ectomicrobiota, el género dominante fue *Pseudomonas*, seguido de *Yersinia* y *Serratia* por la abundancia de sus secuencias y en menor grado algunos grupos de bacterias no clasificados de la familia Enterobacteriaceae. Por su parte, bacterias no clasificadas de la clase Gammaproteobacteria fueron las más abundantes en la endomicrobiota, seguidas de OTUs clasificadas como parte de la familia Enterobacteriaceae. A partir del estudio de las galerías, se destacaron los géneros *Pseudoxanthomonas* y a miembros de la familia Enterobacteriaceae por su abundancia predominante, seguidos de otros ejemplos como *Pseudomonas*, *Serratia*, *Acidobacteria* Gp1 y *Lactobacillus* (Durand *et al.*, 2015)

Asimismo, se ha ampliado el estudio de la microbiota en insectos a lo largo de las diferentes etapas de su ciclo de vida. En el caso del género *Dendroctonus*, con la especie *D. rhizophagus* en el artículo de Briones-Romero y colaboradores del 2017. En él se consideraron cinco etapas de desarrollo (larva, pupa, teneral o adulto recién terminada la metamorfosis, adultos pre-emergidos y adultos emergidos), entre las cuales se identificaron cuatro phyla bacterianos principales: Proteobacteria, Firmicutes, Actinobacteria y Bacteroidetes. Se encontraron miembros de los géneros *Rahnella*, *Serratia*, *Pseudomonas* y *Propionibacterium* en todas las etapas de la vida, y los primeros tres géneros, particularmente *Rahnella*, fueron predominantes, lo que sugiere que tiene un papel fundamental dentro del microbioma central intestinal (Briones-Romero *et al.*, 2017). Además, este estudio confirmó que la diversidad bacteriana encontrada varía entre las diferentes etapas del ciclo de vida del escarabajo, tal como ya se había estudiado en la especie china *D. armandi* por medio de métodos dependientes de cultivo (Hu *et al.*, 2013).

Otro enfoque de gran potencial derivado del estudio extensivo de la microbiota, consiste en la predicción de funciones o rutas metabólicas asociadas a los taxa identificados con base en los genes de las comunidades microbianas, y su contraste con la información disponible en bases de datos (White *et al.*, 2016; Mineeva *et al.*, 2020). Uno de los ejemplos más destacables de este tipo de análisis, lo encontramos en el trabajo de Adams y colaboradores del 2013, en el cual se identificaron funciones de la microbiota asociadas con el metabolismo de terpenos en *D. ponderosae*. Esto cobra importancia debido a que en respuesta a los ataques de los escarabajos, las coníferas secretan y liberan terpenos, que son tóxicos para los insectos que se alimentan del floema (Philips *et al.*, 1999). También se ha sugerido que algunas especies de *Pseudomonas* aisladas de *D. ponderosae* y *D. valens* mediante cultivos, facilitan la colonización de los árboles hospederos gracias a su capacidad de metabolizar las moléculas terpenoides (Adams *et al.*, 2011; Adams *et al.*, 2013).

Los microorganismos también pueden estar involucrados en procesos de inmunidad en algunas especies de escarabajos (Hurst *et al.*, 2009). Se ha visto que algunas bacterias simbióticas pueden proteger a los insectos contra los hongos entomopatógenos que actúan como agentes de control biológico naturales o implementados por el humano (Scott *et al.*, 2008; Scarborough *et al.*, 2005). Por ejemplo, especies de *Streptomyces* aisladas de *D. frontalis* producen micangimicina, una molécula antifúngica que inhibe el crecimiento de hongos antagonistas (Scott *et al.*, 2008).

Se ha reportado que las bacterias simbióticas complementan la dieta de insectos floófagos con aminoácidos, vitaminas y nitrógeno (Douglas *et al.*, 2009; Gibson *et al.*, 2003). Entre los grupos relevantes que componen la microbiota de *Dendroctonus* y que han sido estudiadas por medio de métodos dependientes al cultivo, cabe mencionar a *Enterobacter agglomerans* y *Enterobacter* spp. Estas especies fijadoras de nitrógeno están asociadas con larvas y adultos de la especie *D. valens* y *D. rhizophagus* (Morales-Jimenez *et al.*, 2013). También se han aislado géneros del phylum Proteobacterias (*Pseudomonas*, *Acinetobacter* y *Rahnella*) y de Actinobacteria (*Arthrobacter*), provenientes del sistema digestivo de *D. rhizophagus*, las cuales son capaces de descomponer celulosa, xilanos, pectina, almidón, lípidos y ésteres (Morales-Jimenez *et al.*, 2012; Briones Roblero *et al.*, 2016).

### **Fungi: Hongos miceliales y Levaduras**

A lo largo de la subfamilia Scolytinae se encuentran clados con diferente grado de asociación simbiótica hacia los hongos. Aunque las simbiosis estricta se reconozca tanto en los escarabajos ambrosiales de esta subfamilia, caracterizados por su hábito xylomicetófago,

se ha encontrado que la gran mayoría de los escarabajos descortezadores de esta subfamilia no presentan una simbiosis estricta, al haberse reportado descendencia que puede sobrevivir sin los hongos que inocula la madre en condiciones de laboratorio (Six *et al.*, 1998). Algunos otros integrantes de la subfamilia Scolytinae con diferentes hábitos alimenticios (espermófagos o dependientes de la médula arbórea) también pueden presentar simbiosis fúngica, sin embargo se sabe menos sobre géneros (Six, 2012).

En el caso del género *Dendroctonus*, la diversidad y el papel que desempeñan los hongos simbiotes no se comprende completamente en todas sus especies, no obstante, las especies de hongos ascomicetos del orden Ophiostomatales se han caracterizado como el componente principal de estas comunidades de hongos simbiotes, particularmente los géneros de teleomorfos *Ophiostoma*, *Ceratocystis* y *Grosmannia*. También se ha registrado de manera recurrente la presencia de otro hongo ascomiceto del género *Ceratocystis* (orden Microascales), tanto en las galerías de los escarabajos del género *Dendroctonus*, como en algunas comunidades micangiales (Six, 2012). Además, algunas de las especies pertenecientes a los clados con micangio protorácico o maxilar parecen estar constantemente asociadas con basidiomycetes, particularmente con los géneros *Entomocorticium* y en menor medida con *Phlebiopsis* (Six, 2012; Hisau y Harrington, 2003). Dichos simbiotes están asociados con beneficios a nivel ecosistémico, ya que promueven la descomposición de la corteza de los árboles que infectan y contribuyen a la nutrición de larvas, incluso de especies que no se relacionan directamente con el hongo (Six, 2012; Hisau y Harrington, 2003; Castello, 1976).

Entre los integrantes del género *Dendroctonus* se ha observado que la simbiosis con hongos, tiene un papel importante en su desarrollo y supervivencia, aunque este efecto puede variar dependiendo de la especie de hongo presente. Por ejemplo, para *D. ponderosae*, el ascomiceto *Grosmannia clavigera* permite el desarrollo más rápido de la cría y una mayor producción de huevos que *Ophiostoma montium* (Six y Paine, 1998). En el caso de *D. frontalis*, el basidiomiceto *Entomocorticium* sp. influye en una mayor supervivencia del huésped, un tamaño corporal más grande y una mayor acumulación de lípidos en los tejidos del escarabajo (Bridges, 1983; Coppedge *et al.*, 1995; Six ). Incluso se ha observado que la presencia de tinciones características de *Ophiostoma minus* en el tejido arbóreo, puede tener un efecto antagónico en la construcción de las cámaras de *D. frontalis*, el cual tiende a girar para evitar alimentarse en estas áreas (Nevill y Safranyik, 1996; Six, 2012).

Tal como las bacterias y los hongos micangiales (de crecimiento micelial o filamentoso), se han encontrado levaduras (hongos unicelulares) simbiotes que tienen un papel importante en la

nutrición de insectos, usualmente debido a sus productos metabólicos y su influencia en procesos digestivos, de desintoxicación y producción de feromonas (Douglas, 2009). De igual manera, se han aislado géneros de levaduras pertenecientes a Ascomycota (*Candida*, *Cyberlindnera* y *Zygoascus*) y a Basidiomycota (*Rhodotorula*) a partir del sistema digestivo de *D. rhizophagus*, las cuales son capaces de degradar compuestos vegetales complejos como la celulosa, el almidón y los lípidos (Briones Roblero *et al.*, 2016).

Frecuentemente se ha reportado la presencia de levaduras a partir de muestras de las galerías de *Dendroctonus* y de otros géneros de escarabajos descortezadores (Hofstetter *et al.*, 2015). En particular, se han aislado levaduras de las galerías y de todas las etapas del ciclo de vida de ejemplares de ambos sexos de *D. pseudotsugae* y de *D. micans* (Lewinsohn *et al.*, 1994; Bridges *et al.*, 1984). Además, diferentes especies han sido identificadas como integrantes de la comunidad de hongos micangiales de *D. brevicomis* (Davis y Hofstatter, 2011) y de *D. ponderosae*, especie en la cual también se aislaron levaduras a partir del exoesqueleto de los escarabajos (Six, 2003). La mayoría de las levaduras que se han logrado identificar en el género pertenecen a la familia Saccharomycetaceae (Div. Ascomycota), destacando a géneros como *Ogastea* (= *Pichia*), *Hansenula*, *Kuraishia* y *Candida*. Al comparar las levaduras encontradas en diez especies del género se encontró que algunas levaduras parecen estar asociadas preferentemente con escarabajos descortezadores que se alimentan de diferentes géneros de coníferas y, por lo tanto, se ha propuesto que las defensas del árbol hospedero pueden limitar la diversidad de la comunidad de levaduras en *Dendroctonus*. (Rivera *et al.*, 2019)

Algunas funciones que se han relacionado con las levaduras simbiotes de *Dendroctonus* incluyen a la atracción o repulsión de escarabajos de otras especies, depredadores y parásitos a partir de las emisiones gaseosas generadas por de cultivos de levaduras o bien, el metabolismo de feromonas de agregación producidas por *D. ponderosae* (cis y trans-verbenol) en verbenona, la cuál tiene un efecto contrario al inhibir la agregación de escarabajos de esta especie (Hunt y Borden *et al.*, 1990; Boone *et al.*, 2008). Además, algunas levaduras de los géneros simbiotes aislados son capaces de sobrevivir y metabolizar compuestos terpenoides inhibidores del crecimiento de hongos micangiales, por lo que la presencia de estas levaduras puede ser importante para la sobrevivencia de la microbiota en conjunto y para potenciar la tolerancia del escarabajo a las fitotoxinas (Davis, 2011). Adicionalmente se ha encontrado que la interacción de las levaduras simbiotes con otros hongos, frecuentemente asociados a este nicho, puede tener diferentes efectos como el aumento de la tasa de crecimiento del basidiomiceto *Entomocorticium* sp. ante las emisiones volátiles de las levaduras o la inhibición del crecimiento del hongo entomopatógeno *Beauveria bassiana* (Davis y Hofstetter, 2011).

## Nemátodos

El phylum Nematoda se caracteriza por su cuerpo vermiforme largo, delgado y por ser uno de los más diversos y numerosos entre los metazoarios. Suelen ser microscópicos, pero pueden alcanzar tamaños de hasta 8m. En general, es un grupo que se encuentra sub descrito y sub estudiado, a excepción de la especie modelo *Caenorhabditis elegans*, de gran importancia en campos como la neurobiología y la biología del desarrollo (Brusca *et al.*, 2016). Se estima que existen entre 40,000 y 500,000 especies de nematodos asociadas a insectos, muchas de las cuales han sido frecuentemente asociadas con los escarabajos descortezadores (Giblin-Davis *et al.*, 2013).

Se han reportado interacciones de tipo comensal, mutualista, parásito o parasitoide en nemátodos simbioses de escarabajos descortezadores. Aunque muchas de estas no han sido estudiadas a profundidad, es posible que formen parte de una simbiosis multipartita en la que las relaciones positivas o negativas están sujetas a cambios dependientes del sistema y de sus condiciones ecológicas (Hofstatter, 2015; Relman, 2008). Algunos nemátodos parásitos pueden matar a su hospedero, por lo que han sido sugeridos como potenciales recursos para el control de las poblaciones de diferentes especies de escarabajos descortezadores (Hofstatter, 2015).

Las especies de *Dendroctonus* en las que se ha estudiado a la diversidad de nematodos simbioses incluyen a *D. adjunctus*, destacada por su capacidad de albergar más de 25 especies de nemátodos (Massey, 1966) y *D. pseudotsugae*, en donde se reportó una reducción de la fecundidad de sus hembras al ser infectadas por el nematodo parasitoide *Contortylenchus reversus* (Thong y Webser, 1975). También se ha encontrado al nemátodo endocomensal *Rhabditis obtusa* dentro de los intestinos de *D. rufipennis*, *D. ponderosae* y *D. adjunctus* (Massey, 1956; Massey, 1966).

En el caso de *D. rufipennis*, se ha identificado el papel de las especies de nemátodos simbioses como parte de un ecosistema microbiano más grande, en el cual se ha sugerido que los nemátodos se alimentan de los hongos simbioses y mantienen una interacción comensal con el escarabajo (Cardoza *et al.*, 2008). Además se ha descrito una relación de exclusión entre poblaciones de nemátodos simbioses de este escarabajo debido a que la presencia de *Contortylenchus reversus* previene infecciones por el parásito *Sphaerulariopsis dendroctoni* (Thong y Webster, 1983).

## Métodos de estudio de la microbiota

Debido a que un gran porcentaje de las especies que conforman la microbiota no son aislables en un medio de cultivo, estas interacciones han sido estudiadas por medio del conjunto de genes recuperados de muestras de la comunidad, conocidos como *microbioma*. Esto ha sido posible gracias

al desarrollo de aproximaciones tecnológicas con las que se han evaluado las abundancias relativas y los niveles de actividad génica de los microbiomas, ya sea a partir de la extracción y secuenciación inespecífica de todo el material genómico, o de la secuenciación dependiente de primers asociados a regiones “barcode” específicas para determinados grupos taxonómicos, por ejemplo ITS 1 y 2 en hongos o CO1 en animales (Tanabe y Toju, 2013; Aschenbrenner *et al.*, 2016).

La investigación por medio de técnicas de secuenciación masiva en paralelo, ha permitido que la cantidad de información disponible sobre comunidades microbianas como parte integral del ecosistema o de sus hospederos multicelulares, haya aumentado de manera acelerada durante la última década (White *et al.*, 2016). No obstante, el empleo de estas técnicas depende de un proceso de limpieza, ensamble y análisis computacional de datos que no ha sido estandarizado dentro de la *metagenómica*. Esta rama de las disciplinas “ómicas”, está comprendida por aquellas técnicas que recuperan toda la información genómica disponible de múltiples taxa a partir de una muestra de DNA ambiental o bien, proveniente de una comunidad huésped, por medio de técnicas de SMP. Por ejemplo, la obtención del microbioma intestinal o las posibles inferencias sobre la dieta del hospedero a partir de heces (Elisen, 2007). Al igual que en muchas otras disciplinas, la diversidad de ensayos, métodos de análisis y datos públicos disponibles ha generado una gran variedad de herramientas útiles para diferentes etapas del análisis y para los fines de cada proyecto, por ejemplo, la limpieza y el ensamble de los datos, asignación taxonómica, la predicción de rutas metabólicas y el modelado de interacciones microbianas dentro de la comunidad (Segata *et al.*, 2014; Mineeva *et al.*, 2020).

Los métodos de asignación taxonómica han continuado desarrollándose a la par que los métodos para la obtención de la información genética asociada a la microbiota. El flujo de trabajo bioinformático generalmente incluye pasos de filtrado por calidad, de corrección de errores o de “ruido” de secuenciación y de eliminación de reads o lecturas quiméricas. Al igual que con otros métodos de estudio de la microbiota mediante SMP, al analizar amplicones “barcode”, se agrupan a las lecturas filtradas basándose en la similitud de las secuencias para finalmente realizar la clasificación de estas OTUs (Brietawaser *et al.*, 2019). Una ventaja de esta aproximación es que algunas bases de datos de genes “barcode” como los genes 16S y 18S rRNA contienen genes de millones de especies. Sin embargo, las bases de datos genómicas sólo contienen algunos cientos de miles de especies y a menudo manejan un sistema de clasificación menos funcional para este tipo de trabajos (DeSantis *et al.*, 2006; Cole *et al.*, 2005; Carlton *et al.*, 2002).

Un enfoque alternativo para la asignación taxonómica, se basa en la clasificación de lecturas de amplicones antes de agruparlos en OTUs por medio de métodos de asignación de información metagenómica. Múltiples estrategias han sido desarrolladas para realizar esto dependiendo de los

objetivos del trabajo, entre las cuales incluye el alineamiento directo de lecturas, el mapeo de k-mers, el ensamble de las lecturas en metagenomas o secuencias contiguas de mayor tamaño (contigs), la asignación por medio de genomas microbianos completos, el alineamiento de genes marcadores o la traducción de DNA y el subsecuente alineamiento de secuencias proteicas (Brietawaser *et al.*, 2019).

Si bien, anteriormente la mejor estrategia para la asignación taxonómica era utilizar la herramienta BLAST, Basic Local Alignment Search Tool (Altschul *et. al.*, 1990) para comparar cada lectura con todas las secuencias disponibles en bases de datos exhaustivas, a medida que las bases de datos de referencia y el tamaño de los datos de secuenciación han crecido, la asignación taxonómica mediante BLAST se ha vuelto computacionalmente inviable, lo que ha llevado al desarrollo de múltiples clasificadores metagenómicos mucho más rápidos, aunque con menor sensibilidad que BLAST (Brietawaser *et al.*, 2019).

La precisión y confiabilidad de los métodos de asignación taxonómica dependen de la base de datos de información genómica o metagenómica empleada como referencia, así como de la taxonomía manejada por ésta. Las bases de datos del Centro Nacional para la Información Biotecnológica de EEUU (NCBI), en conjunto con el Laboratorio Europeo de Biología Molecular (EMBL) y el Banco de Datos de DNA de Japón (DDBJ), que en conjunto forman la Colaboración Internacional de Bases de Datos de Secuenciación de Nucleótidos (INSDC), comparten el sistema de nomenclatura estándar y de árbol taxonómico jerárquico (Cochrane *et al.*, 2016). Por su parte, las bases de datos más modernas, a menudo creadas para ser empleadas por clasificadores metagenómicos, como Greengenes (DeSantis *et al.*, 2006), RDP (Cole *et al.*, 2005) y SILVA (Quast *et al.*, 2012), manejan su propia taxonomía para evitar problemas derivados del sistema jerárquico. Este sistema fue diseñado para el estudio de organismos multicelulares como plantas y animales (Balvočiūtė, et al., 2017), por lo que se ha enfrentado a diferentes desafíos al aplicarlo en el estudio de otros grupos, como la presencia de diferentes niveles de similitud entre secuencias en un mismo nivel taxonómico en bacterias o arqueas (Rosselló-Mora y Amann, 2015), la homonimia que presentan las especies fúngicas con formas teleomórficas (sexuadas) y anamórficas (asexuadas) asociadas a diferentes códigos de identificación taxonómica (taxID) (Taylor, 2011), o el rezago de la asignación taxonómica de la diversidad microbiana no cultivable bajo nombres *Candidatus* (Konstantinidis y Rosselló-Mora., 2015).

La creciente capacidad de obtener información sobre las comunidades microbianas, ha permitido reevaluar las estimaciones sobre la diversidad. Desde el siglo XIX, en el que se consideraba a los hongos como parte del reino vegetal, se estimaba que este sería el grupo más diverso de aquella clasificación (Hawksworth y Lücking, 2017). Actualmente se han descrito al menos 120,000 especies

de hongos. Sin embargo, las predicciones sobre la microdiversidad global que consideran las bases de datos metagenómicas y emplean los métodos de ajuste de escalas, han estimado entre 100 mil millones ( $10^{11}$ ) y 1 billón ( $10^{12}$ ) de especies microbianas distribuidas globalmente, de las cuales se cree que al menos 1% pertenece al reino Fungi (Locey y Lennon, 2016). Los rangos más recientemente aceptados para este reino, rondan entre 2.2 y 3.8 millones de especies (Hawksworth *et al.*, 2017; Cuadros-Orellana *et al.*, 2013). La identificación y el estudio de esta microdiversidad cobra especial importancia en el caso de los simbioses de artrópodos, ya que las rutas metabólicas microbianas están asociadas con la alta proporción C:N de la dieta de los artrópodos y contribuyen significativamente en la fijación de nitrógeno en ecosistemas terrestres, o bien, han influido de manera importante en la diversificación de algunos grupos de artrópodos (Nardi *et al.*, 2002).

El reciente desarrollo de los estudios sobre las comunidades microbianas asociadas a insectos por medio de técnicas de SMP, puede ser ejemplificado en el género *Dendroctonus*, en el que ya se han realizado múltiples estudios sobre la microbiota bacteriana intestinal por medio del gen ribosomal 16S (Hernandez-García *et al.*, 2017; Briones-Robiero *et al.*, 2017; Hu *et al.*, 2015), e incluso comparaciones de la microbiota bacteriana interna con la comunidad asociada a la cutícula (Durand, 2015). El presente estudio aspira a complementar el acervo de conocimiento sobre esta interacción, por medio del estudio de la diversidad microbiológica asociada a múltiples especies del género *Dendroctonus*.

## **Métodos de secuenciación de representación reducida del genoma**

El estudio del microbioma a menudo requiere del diseño de primers y de capacidad económica, por lo que el desarrollo de métodos alternativos como el uso de información previamente generada con técnicas de secuenciación de representación reducida del genoma (SRRG), representa un área de oportunidad para plantear nuevas preguntas de investigación (Holmes y Rabosky, 2018; Orantes *et al.*, 2018).

Las técnicas de SMP permiten recuperar información representativa del genoma a través de distintos protocolos de laboratorio como la genotipificación por secuenciación (GBS), la secuenciación asociada a sitios de restricción (RADseq) y sus variantes (ddRAD, 3RAD). Estas técnicas están fundamentadas en la fragmentación del genoma por medio de una o más enzimas de restricción, para generar secuencias de menor tamaño compartidas entre las diferentes muestras. La



variación encontrada en los loci ensamblados, permite realizar inferencias filogenéticas o sobre la estructura genética de la población (Andrews, 2016, Schweyen *et al.*, 2014).

Se ha propuesto que modificando el procesamiento bioinformático de los datos generados a partir de técnicas de SRRG, es posible recuperar información del microbioma asociado a los organismos, debido a que estas técnicas generan lecturas de manera aleatoria (Andrews, 2016). Además de las ventajas asociadas al bajo costo y a la versatilidad de estas técnicas, se cree que pueden ser empleadas como métodos exploratorios y posibles herramientas para identificar linajes de la microbiota por medio de enfoques independientes de primers o de *metabarcoding* (Holmes y Rabosky, 2018).

Anteriormente se ha evaluado la capacidad de recuperar información genómica de comunidades bacterianas ya caracterizadas en poblaciones de vertebrados a partir de datos de shotgun no mapeados o de datos de RADSeq (Holmes y Rabosky, 2018; Hooper, 2018), sin embargo no se han realizado sondeos de la microbiota que incluyan múltiples especies de un género hospedero en particular. El desarrollo de este método en niveles taxonómicos supraespecíficos, representa un paso importante para resolver preguntas de importancia evolutiva, así como para explorar a la microbiota en especies menos estudiadas de géneros propuestos como modelos de estudio de la simbiosis, como *Dendroctonus*.

En el presente trabajo se utilizaron los datos de RADseq generados por Godefroid y colaboradores (2019) para caracterizar la diversidad microbiológica dentro del género de escarabajos descortezadores, *Dendroctonus*. Para esto se generaron ensamblados usando múltiples métodos disponibles en el pipeline ipyrad v.07.28 (Eaton, 2014), y se evaluó la cantidad de información del microbioma recuperada al variar el parámetro de umbral de agrupamiento o Clustering Threshold (CT) y *min depth majority rule*, que corresponde a la profundidad mínima que debe tener una base para ser llamada al ensamblado. La determinación taxonómica de las secuencias se realizó por medio de un análisis de BLAST, lo cual permitió procesar la información en R para realizar inferencias sobre la relación entre los microorganismos encontrados y las especies de escarabajo hospederos.

## JUSTIFICACIÓN

El análisis bioinformático de los datos de RADseq previamente publicados, permitirá evaluar la capacidad de estudiar a las comunidades microbianas asociadas a un hospedero mediante técnicas de representación reducida del genoma. La comparación de los resultados de este trabajo con respecto a la diversidad de especies encontrada en otros estudios del microbioma de escarabajos

descortezadores, realizados por medio de enfoques tradicionales, permitirá emplear este trabajo como referencia para futuras investigaciones sobre métodos de estudio de la diversidad microbiana en escarabajos descortezadores.

Además, el estudio extendido de la microbiota asociada a un género altamente representado en México, como *Dendroctonus*, permite aumentar el conocimiento de la diversidad microbiológica a menudo subestimada como un componente del capital biológico nacional, ya que 13 de las 21 especies del género de escarabajos se han encontrado en el territorio nacional y una de ellas, *D. rhizophagus*, es endémica de la Sierra Madre Occidental (Armendáriz-Toledano *et al.*, 2018). La determinación de los taxa que mantienen una potencial simbiosis con este género de escarabajos, es un parteaguas para enfocar estudios en especies cuya microbiota tiene un impacto socioambiental prioritario.

## **OBJETIVOS**

Caracterizar la diversidad de microorganismos asociados a especies del género *Dendroctonus* mediante el análisis de secuencias de RADseq previamente publicadas.

### **Objetivos particulares**

Evaluar la capacidad de recuperar información metagenómica de la microbiota de insectos del género *Dendroctonus* a partir de secuencias obtenidas mediante la técnica de RADseq.

Emitir recomendaciones para optimizar la obtención de información genómica de la microbiota de insectos a partir de secuencias previamente publicadas.

Comparar la diversidad microbiana encontrada en las diferentes especies del género *Dendroctonus* con estudios previos sobre la diversidad de determinados grupos de importancia biológica asociados a estos escarabajos.

Identificar patrones de asociación de determinados taxa microbianos entre las diferentes especies de *Dendroctonus* para proponer posibles relaciones cofilogenéticas o ecológicas.

## MÉTODO

### Obtención de datos previamente publicados

Los datos de secuenciación de RADseq del género *Dendroctonus* provienen de la publicación de Godefroid *et al* de 2019, en la cual se realizó una reconstrucción filogenética incluyendo la mayor parte de las especies descritas en este género, exceptuando a *D. mesoamericanus*, *D. vitei* y *D. parallellocollis*.

La extracción de ADN se realizó mediante una incisión entre el tórax y el abdomen de los especímenes para posteriormente realizar la digestión enzimática de los escarabajos completos. La secuenciación se llevó a cabo empleando la técnica de RADseq con una sola enzima, *PstI*. Se prepararon librerías genómicas siguiendo el método de Baird *et al.* (2008) y de Etter *et al.* (2011) para llevar a cabo la secuenciación de ambos extremos en MGX-Montpellier GenomiX para 2 \* 125nt en una plataforma de Illumina HiSeq 2500. El procesamiento de los datos se realizó con el pipeline “RADIS-Stacks” (Cruaud, 2016), con el que se ensamblaron 23,570 tags de RAD, permitiendo 12 mismatches entre las secuencias para formar un locus individual y que cada locus estuviera compartido en al menos 35% de las muestras, para considerarlo en la matriz final empleada en el artículo (Godefroid *et al.*, 2019). Los datos crudos de este proyecto se compartieron sin barcodes bajo el código de acceso en la base de datos del NCBI: PRJNA530572.

Para evaluar la capacidad de recuperar información genética de comunidades de microorganismos asociados a los escarabajos, empleamos el pipeline ipyrad v.0.7.28 (Eaton, 2014), cuya función es análoga a la del programa empleado en el artículo original para procesar las secuencias crudas e identificar aquellas que son homólogas entre las muestras. El uso más extendido de ipyrad consiste en ensamblar un conjunto de datos de RADSeq bajo una serie de parámetros definidos. Los parámetros más destacados para este proyecto fueron el porcentaje de similitud para realizar el agrupamiento entre secuencias o CT, el número mínimo de muestras en las que se debe compartir un locus para incluirlo en la matriz final o *min\_sample\_locus* y el valor de profundidad que debe tener una base para ser recuperada por la regla de la mayoría o *mindepth\_majrule*.

El flujo de trabajo de este software ha sido dividido por sus creadores en siete pasos: en el paso 1 el programa carga y asigna los datos a cada muestra, en el paso 2 realiza un filtrado por calidad basado en los valores de Probabilidad de factor de error de base (Phred, por sus siglas en inglés) de los archivos “fastq”, en el paso 3 cuantifica el número de lecturas únicas y las agrupa por similitud dentro de cada muestra. El paso 4 calcula el porcentaje de heterocigosidad y de error con los cuales en el paso 5, el programa determina las secuencias de los alelos consenso, llama a las bases con profundidad suficiente y filtra los loci con base en número máximo de sitios indeterminados (N’s). En

el paso 6 el programa identifica y agrupa loci ortólogos entre las diferentes muestras). Finalmente en el paso 7, el programa permite aplicar diferentes filtros a los loci ortólogos con base en una serie de parámetros definidos por el usuario y genera archivos de salida en diversos formatos para su aprovechamiento en análisis posteriores (Eaton y Overcast, 2020).

El software ipyrad ofrece tres aproximaciones diferentes para el ensamble de loci: ensamble *de novo*, *con referencia* y *de novo menos referencia*. El ensamble *de novo* mapea los fragmentos aleatorios del genoma generando clusters mediante el uso de un algoritmo de agrupación global de alineamiento, recurriendo al programa VSEARCH (Rognes, 2016), el cual permite la incorporación de cierto grado de inserciones y deleciones al identificar homologías. En cambio, el ensamble *con referencia* requiere de una secuencia proporcionada por el usuario para mapear los fragmentos por medio de los programas bwa y bedtools (Li y Durbin, 2009; Quinlan y Hall, 2010), este algoritmo identifica homologías y descarta todo lo que no se parece a la referencia. Por último, el ensamble *de novo menos referencia* elimina aquellas lecturas que coinciden con una secuencia de referencia y ensambla *de novo* aquellas restantes. (Eaton y Overcast, 2020).

## **Ensamble de datos**

El primer enfoque empleado para explorar los datos y evaluar los recursos de cómputo necesarios para recuperar loci microbianos, consistió en realizar dos ensambles *de novo* con el set completo de datos de RADSeq: uno con un valor de *clustering\_threshold* (CT) de 0.97 (denominado *Denovo\_97*) y otro con un valor de 0.95 (denominado *Denovo\_95*).

Este parámetro corresponde al porcentaje de similitud al que dos lecturas se consideran homólogas. A pesar de la variabilidad asociada a los datos de RADSeq, se ha demostrado por medio de métodos empíricos que el impacto de este parámetro tiende a aumentar al adoptar valores altos, de más de 0.96, fijando a esta cifra como un umbral alrededor del cual suele ser óptimo iniciar a evaluar su efecto en la cantidad de datos recuperados (McCartney-Melstad *et al.*, 2019). Cuando se emplea un valor de CT muy bajo, la probabilidad de agrupar loci parálogos en una misma muestra o entre muestras diferentes aumenta, o bien, si se elige un valor muy alto, el algoritmo puede excluir a aquellos loci homólogos que presentan alta variabilidad genética entre sí. Además, para cumplir con los objetivos de este estudio se priorizaron las pruebas con valores altos de homología entre loci, considerando la influencia del azar en esta técnica de secuenciación y de la baja concentración de DNA microbiano.

Posteriormente, se evaluó el desempeño de los ensamblados *de novo* - referencia en el set de datos completo, utilizando a la única secuencia del genoma completo del género, de *D. ponderosae*, como referencia (NCBI BioProject: PRJNA162621, Keeling *et al.*, 2013) con el fin de filtrar la información genómica de los escarabajos y recuperar únicamente los loci microbianos. Se realizaron pruebas etiquetadas como *DRef\_* con un valor de CT de 0.97 y de 0.95. Al seleccionar a la secuencia del genoma completo (draft) se optó por realizar los ensamblados con el genoma de las hembras de *D. ponderosae*, debido a la afinidad que tienen algunos simbioses con las hembras de otras especies de *Dendroctonus*.

Después se buscó optimizar el desempeño de esta aproximación ensamblando únicamente los reads de las muestras correspondientes a la especie de la cual se recuperó la secuencia de referencia, *D. ponderosae* (codificadas como SRR8838372, SRR8838376, SRR8838378, y SRR8838386 dentro los archivos SRA del bioproyecto). Se esperaba que al limitar el ensamble a las muestras de *D. ponderosae* usando una secuencia de referencia de la misma especie, se filtraría una mayor cantidad de loci del hospedero y por ende se recuperarían taxa microbianos de manera más eficiente. Esto se evaluó mediante dos pruebas etiquetadas como *DPond\_* con valores de CT de 0.97 y de 0.95.

Adicionalmente, se realizó el procedimiento de determinación taxonómica de la secuencia del genoma completo de *D. ponderosae* que sirvió como referencia en los ensamblados *de novo* - ref. Este proceso se llevó a cabo filtrando los taxa de la base de datos del NCBI que emplea BLAST como referencia en *Miztli*, para limitar la búsqueda a la información genómica del Reino Fungi que podía presentarse en la secuencia del genoma de referencia *a priori*. Esta posibilidad cobró importancia entre las consideraciones del proyecto debido al método de ensamble del programa ipyrad *de novo* - referencia, el cual, de realizarse con una secuencia genómica de referencia contaminada por información del microbioma, puede causar la pérdida de una porción de la diversidad microbiana resultando contraproducente para los objetivos de este trabajo. Esta búsqueda puede replicarse para limitar la base de datos a secuencias bacterianas o de otros taxa determinados, con el fin de evaluar la posible influencia de contaminación microbiana o bien, la presencia de información genómica de un simbionte reportado en un genoma de referencia.

La siguiente aproximación para explorar los diferentes métodos de recuperación de loci microbianos, consistió en realizar dos ensamblados + referencia utilizando como referencia al genoma completo de uno de los hongos simbioses característicos de algunas especies de *Dendroctonus*, *Grosmannia clavigera* (NCBI BioProject: PRJNA264104, DiGustini *et al.*, 2011). Se propuso este método por el potencial que tiene de ensamblar una mayor cantidad de loci fúngicos, debido a la

simbiosis micangial de este género de hongos. Estas pruebas fueron etiquetadas como *GRef\_* y se realizaron con los dos valores de CT base de este proyecto: 0.97 y 0.95.

Al observar la baja cantidad de loci microbianos recuperados por estos primeros ensamblajes *+referencia*, se revisaron los archivos de estadísticas (.stats) de resumen que arroja ipyrad (.stats), en los cuales se desglosa la cantidad de loci filtrados a lo largo de un ensamblaje. Al encontrar una gran cantidad de loci excluidos del ensamblaje debido a su baja profundidad de secuenciación, (Anexo 1), se llevó a cabo un nuevo ensamblaje con un valor de CT de 0.97 y con el parámetro *mindepth\_majrule* disminuido de 6 a 4 para evaluar con ello el potencial de recuperar más loci microbianos a partir de lecturas con menor *profundidad*. Dicha propiedad es definida como la media del número de veces que cada base de un genoma, o metagenoma, tiene una lectura que se alinea en esa posición. Tras no hallar los resultados esperados, se disminuyó el valor de este parámetro al mínimo recomendado, 2. Estos últimos dos ensamblajes fueron nombrados *GRef\_97\_m4* y *GRef\_97\_m2*.

Los métodos de consenso por *majority rule* son empleados para asegurar que todas las bases recuperadas de los reads crudos tengan una profundidad mayor que un porcentaje estadístico determinado, el cual está fijado en 6 en el parámetro *mindepth\_statistical* por default. El método de “majority rule calling” fue desarrollado para optimizar estudios filogenéticos (Jermin *et al.*, 1997), sin embargo ha demostrado tener una gran utilidad en el ensamblaje de secuencias de RADSeq para tratar con datos en los que muchos loci se excluyen debido a una baja profundidad de secuenciación (Eaton y Overcast, 2020).

Finalmente, se evaluó el efecto de reducir el CT a 90 y 85 en el ensamblaje de loci microbianos *GRef\_97\_m2*, para confirmar que se mantenía la eficiencia de manera proporcional a la disminución del CT tal como se describió en el trabajo de Holmes y Rabosky del 2018. Estos últimos ensamblajes fueron etiquetados como *GRef\_90\_m2* y *GRef\_85\_m2*.

<b>Tipo de ensamblaje</b>	<b>Nombre del ensamblaje</b>	<b>Clustering threshold (CT)</b>	<b>Datos de RAD</b>	<b>Genoma de Referencia</b>
Denovo	1.- Denovo_97	0.97	Set completo	-
	2.- Denovo_95	0.95		
	3.- DRef_97	0.97		<i>D. ponderosae</i>
	4.- DRef_95	0.95		

Denovo - Referencia	5.- DPond_97	0.97		<i>D. ponderosae</i>
	6.- DPond_95	0.95		
+ Referencia	7.- GRef_97	0.97	Set completo	<i>G. clavigera</i>
	8.- GRef_95	0.95		
	9.- GRef_97_m4	0.97 *majrule = 4		
	10.- GRef_97_m2	0.97 *majrule = 2		
	11.- GRef_90_m2	0.90 *majrule = 2		
	12.- GRef_85_m2	0.85 *majrule = 2		

**Tabla 1.** Tipos de ensamblajes realizados con ipyrad. Los ensamblajes fueron denominados de acuerdo al método y al valor de CT empleado, en caso de haber modificado también el parámetro *mindepth\_majorityrule*, se agregó la etiqueta “m” y su valor al final del nombre.

En todos los ensamblajes se modificó el parámetro *min\_samples\_locus* de 4 a 1. Este valor corresponde al número mínimo de muestras en las que debe estar compartido un locus para ser incluido en la matriz final del ensamblaje. El aumento en el valor de este parámetro está relacionado con la disminución de loci recuperados en el ensamblaje final y con el aumento de datos faltantes. El valor estándar del parámetro en el pipeline corresponde a 4, ya que este es el número mínimo de muestras recomendado para incluir un loci en una reconstrucción filogenética, no obstante, en el presente trabajo se le asignó un valor mínimo de 1 para evitar la pérdida de información por falta de representatividad y de esta manera priorizar la caracterización de la diversidad.

Debido al alto poder de cómputo requerido para algunos análisis, en particular los de las categorías *Denovo* y *DRef*, los primeros ensamblajes con este método se realizaron como prueban el cluster de supercómputo de la Comisión Nacional para el Conocimiento y Uso de la Biodiversidad

(CONABIO). Posteriormente todas los ensamblés se repitieron en el cluster de supercómputo de la UNAM, Miztli como parte del proyecto “Análisis bioinformático de datos a escala genómica para el estudio de la evolución de grupos de insectos neotropicales FASE III” del Dr. Alejandro Zaldivar Riverón (Instituto de Biología).

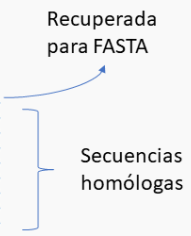
## Determinación taxonómica y análisis

Para cada matriz obtenida por los diferentes ensamblés se recuperó el archivo *.loci* generado por ipyrad, el cual presenta los loci enlistados en grupos, con el nombre de la muestra de donde proviene la secuencia incluido antes de cada una y con una línea de diagonales separando a los loci diferentes dentro del archivo (Fig. 4). Se empleó una modificación del script de R escrito por Holmes y Rabosky en su publicación de 2018 (Anexo 2), para tomar la primera secuencia de cada locus y reescribirla en archivos con el formato *fasta* necesario para la búsqueda mediante la herramienta BLAST del NCBI (Altschul *et al.*, 1990).

```

//
SRR8838326.sra AAAACGAAATGTATAAACTCACCTCGTTATTTGTTTCAGGATTAAGCCCTGTATAGAGCCTTAAAGACGATTAGAAATAAAGCTCAACTATTA
SRR8838327.sra AAAACGAAATGTATAAACTCACCTCGTTATTTGTTTCAGGATTAAGCCCTGTATAGAGCCTTAAAGACGATTAGAAATAAAGCTCAACTATTA
SRR8838328.sra AAAACGAAATGTATAAACTCACCTCGTTATTTGTTTCAGGATTAAGCCCTGTATAGAGCCTTAAAGACGATTAGAAATAAAGCTCAACTATTA
SRR8838374.sra AAAACGAAATGTATAAACTCACCTCGTTATTTGTTTCAGGATTAAGCCCTGTATAGAGCCTTAAAGACGATTAGAAATAAAGCTCAACTATTA
//
SRR8838326.sra TGGAAGCTATGAATTTTGAACCTTAAGTTTATTCAACGCAAGAACAAAAATGCCGCTGAATTCGAAATTTTTGTTAAATTAAGCTCAAAATGG
SRR8838327.sra TGGAAGCTATGAATTTTGAACCTTAAGTTTATTCAACGCAAGAACAAAAATGCCGCTGAATTCGAAATTTTTGTTAAATTAAGCTCAAAATGG
SRR8838328.sra TGGAAGCTATGAATTTTGAACCTTAAGTTTATTCAACGCAAGAACAAAAATGCCGCTGAATTCGAAATTTTTGTTAAATTAAGCTCAAAATGG
SRR8838374.sra TGGAAGCTATGAATTTTGAACCTTAAGTTTATTCAACGCAAGAACAAAAATGCCGCTGAATTCGAAATTTTTGTTAAATTAAGCTCAAAATGG
//
SRR8838326.sra ACCCGGACACTAAGGCCAGCAAAATCCGATTTCGAAACCGGCTCAAGTTCGAGCTCAAAAATCGGGCATTGGAATGGAAGTCAAGGAGGCCCAATTGAG
SRR8838327.sra ACCCGGACACTAAGGCCAGCAAAATCCGATTTCGAAACCGGCTCAAGTTCGAGCTCAAAAATCGGGCATTGGAATGGAAGTCAAGGAGGCCCAATTGAG
SRR8838328.sra ACCCGGACACTAAGGCCAGCAAAATCCGATTTCGAAACCGGCTCAAGTTCGAGCTCAAAAATCGGGCATTGGAATGGAAGTCAAGGAGGCCCAATTGAG
SRR8838339.sra ACCCGGACACTAAGGCCAGCAAAATCCGATTTCGAAACCGGCTCAAGTTCGAGCTCAAAAATCGGGCATTGGAATGGAAGTCAAGGAGGCCCAATTGAG
SRR8838340.sra ACCCGGACACTAAGGCCAGCAAAATCCGATTTCGAAACCGGCTCAAGTTCGAGCTCAAAAATCGGGCATTGGAATGGAAGTCAAGGAGGCCCAATTGAG
SRR8838342.sra ACCCGGACACTAAGGCCAGCAAAATCCGATTTCGAAACCGGCTCAAGTTCGAGCTCAAAAATCGGGCATTGGAATGGAAGTCAAGGAGGCCCAATTGAG
SRR8838345.sra ACCCGGACACTAAGGCCAGCAAAATCCGATTTCGAAACCGGCTCAAGTTCGAGCTCAAAAATCGGGCATTGGAATGGAAGTCAAGGAGGCCCAATTGAG
SRR8838346.sra ACCCGGACACTAAGGCCAGCAAAATCCGATTTCGAAACCGGCTCAAGTTCGAGCTCAAAAATCGGGCATTGGAATGGAAGTCAAGGAGGCCCAATTGAG
SRR8838374.sra ACCCGGACACTAAGGCCAGCAAAATCCGATTTCGAAACCGGCTCAAGTTCGAGCTCAAAAATCGGGCATTGGAATGGAAGTCAAGGAGGCCCAATTGAG
//
SRR8838326.sra TTTGTTGATCGTATTTTTTTGTGCGGGTATTTGACTTATTTCTTTGCTTTCTATGTTAGTGAGCGAAAATGTTTTGTTCTTTTTATCGTTTCTG
SRR8838328.sra TTTGTTGATCGTATTTTTTTGTGCGGGTATTTGACTTATTTCTTTGCTTTCTATGTTAGTGAGCGAAAATGTTTTGTTCTTTTTATCGTTTCTG
//
SRR8838326.sra GCCGTCGGCTCAGTGATCTACCGGCTCAGAGCTTCTGATACTCTGTTGACTATCCTTTGTTTTTCAGCTGAAAGACTCGATTACTGTGGCTTTGG
SRR8838327.sra GCCGTCGGCTCAGTGATCTACCGGCTCAGAGCTTCTGATACTCTGTTGACTATCCTTTGTTTTTCAGCTGAAAGACTCGATTACTGTGGCTTTGG
SRR8838328.sra GCCGTCGGCTCAGTGATCTACCGGCTCAGAGCTTCTGATACTCTGTTGACTATCCTTTGTTTTTCAGCTGAAAGACTCGATTACTGTGGCTTTGG
SRR8838339.sra GCCGTCGGCTCAGTGATCTACCGGCTCAGAGCTTCTGATACTCTGTTGACTATCCTTTGTTTTTCAGCTGAAAGACTCGATTACTGTGGCTTTGG
SRR8838341.sra GCCGTCGGCTCAGTGATCTACCGGCTCAGAGCTTCTGATACTCTGTTGACTATCCTTTGTTTTTCAGCTGAAAGACTCGATTACTGTGGCTTTGG
SRR8838344.sra GCCGTCGGCTCAGTGATCTACCGGCTCAGAGCTTCTGATACTCTGTTGACTATCCTTTGTTTTTCAGCTGAAAGACTCGATTACTGTGGCTTTGG
SRR8838345.sra GCCGTCGGCTCAGTGATCTACCGGCTCAGAGCTTCTGATACTCTGTTGACTATCCTTTGTTTTTCAGCTGAAAGACTCGATTACTGTGGCTTTGG
SRR8838346.sra GCCGTCGGCTCAGTGATCTACCGGCTCAGAGCTTCTGATACTCTGTTGACTATCCTTTGTTTTTCAGCTGAAAGACTCGATTACTGTGGCTTTGG
SRR8838354.sra GCCGTCGGCTCAGTGATCTACCGGCTCAGAGCTTCTGATACTCTGTTGACTATCCTTTGTTTTTCAGCTGAAAGACTCGATTACTGTGGCTTTGG
SRR8838356.sra GCCGTCGGCTCAGTGATCTACCGGCTCAGAGCTTCTGATACTCTGTTGACTATCCTTTGTTTTTCAGCTGAAAGACTCGATTACTGTGGCTTTGG
SRR8838357.sra GCCGTCGGCTCAGTGATCTACCGGCTCAGAGCTTCTGATACTCTGTTGACTATCCTTTGTTTTTCAGCTGAAAGACTCGATTACTGTGGCTTTGG
SRR8838364.sra GCCGTCGGCTCAGTGATCTACCGGCTCAGAGCTTCTGATACTCTGTTGACTATCCTTTGTTTTTCAGCTGAAAGACTCGATTACTGTGGCTTTGG

```



**Figura 5.** Formato de los archivos *.loci* generados como salida de los ensamblés con ipyrad. Las secuencias homólogas se encuentran agrupadas en loci frente al nombre de la muestra de donde provienen y separadas por dos diagonales. A partir de estos se recupera el de la línea superior para reescribirlos en formato FASTA.

Dada la cantidad de loci obtenidos y el tiempo de trabajo necesario para llevar a cabo los análisis de BLAST en la plataforma en línea del NCBI, estos se realizaron en la supercomputadora Miztli de la UNAM con la versión del software NCBI-BLAST-2.10.0+ (Altschul *et al.*, 1990),



procurando que cada trabajo tuviera asignado un número de cores óptimo. Para todos los procesos se empleó el método *discontiguous megablast*, el cual está optimizado para la comparación de secuencias altamente divergentes, especialmente secuencias de diferentes grupos que tienen alineaciones con bajo grado de identidad (Ma *et al.*, 2002).

Mediante la modificación del script de R desarrollado por Holmes y Rabosky, se recuperaron las secuencias de BLAST que presentaron un porcentaje de similitud mayor al 90% al contrastarlas con la base de datos del NCBI y se concluyó la asignación taxonómica mediante el paquete de R *taxize* (Chamberlain *et al.*, 2020), el cual devuelve archivos con la terminación *.tax*. Se comparó la diversidad de phyla recuperados por cada ensamble filtrando a los taxa considerados como no informativos, tales como aquellos pertenecientes Arthropoda (sobrerrepresentados por secuencias del hospedero, *Dendroctonus*) o aquellos sin un sentido biológico relacionado con la microbiota o con el contexto físico de los escarabajos.

Posteriormente, se empleó este mismo script de Holmes y Rabosky para registrar la información taxonómica hasta nivel de familia de cada match con más de 90% de similitud, con lo que se generaron gráficas con el paquete de R *ggplot2* (Wickham *et al.*, 2016) para visualizar a la diversidad de familias de hongos, plantas y proteobacterias recuperada por los ensambles *Denovo\_97*, *DRef\_97* y *GRef\_85\_m2* debido a la gran cantidad de loci que estos ensambles recuperaron de determinados taxa. En lo particular, se buscaba comparar que los ensambles *Denovo* y *Denovo - Referencia* recuperarán los mismos loci, mientras que *GRef\_m2* destacó por la cantidad de loci relacionados con la secuencia de referencia fúngica que era capaz de recuperar.

Con el fin de describir el tamaño de las secuencias fasta con las que se realizó la asignación taxonómica se utilizó un script original de R para calcular las medidas de tendencia central del tamaño de los loci correspondientes a los diferentes taxa recuperados por los ensambles *Denovo\_97* y *GRef\_85\_m2* (Anexo 3). A partir de esto se realizó una comparación entre el tamaño de determinados taxa de acuerdo a su relación con el contexto biológico del escarabajo en ambos set de datos.

Se generó un script original de R (Anexo 4) que explora los archivos generados a lo largo del flujo de trabajo (*.tax*, *.fasta* y *.loci*), con el fin de identificar y contabilizar las muestras y posteriormente identificar los loci informativos (en su mayoría microbianos, pero también de otros grupos de interés biológico como *Streptophyta*) con más de 90% de similitud con respecto a su coincidencia en la base de datos del NCBI. A partir de este análisis se generaron mapas de calor de la diversidad de phyla en cada especie de *Dendroctonus* para los ensambles *Denovo\_97* y *GRef\_85\_m2*, y de la diversidad de familias de Proteobacteria para *Denovo\_97*.

Se repitió este flujo de trabajo en R para recuperar a los loci con más de 97% de similitud del ensamble *Denovo\_97* y visualizar mediante un mapa de calor a la diversidad de Unidades Taxonómicas Operacionales (OTUs) presente en las diferentes especies de *Dendroctonus*, o bien, de secuencias cuya asignación taxonómica puede llegar al menos hasta el nivel de género.

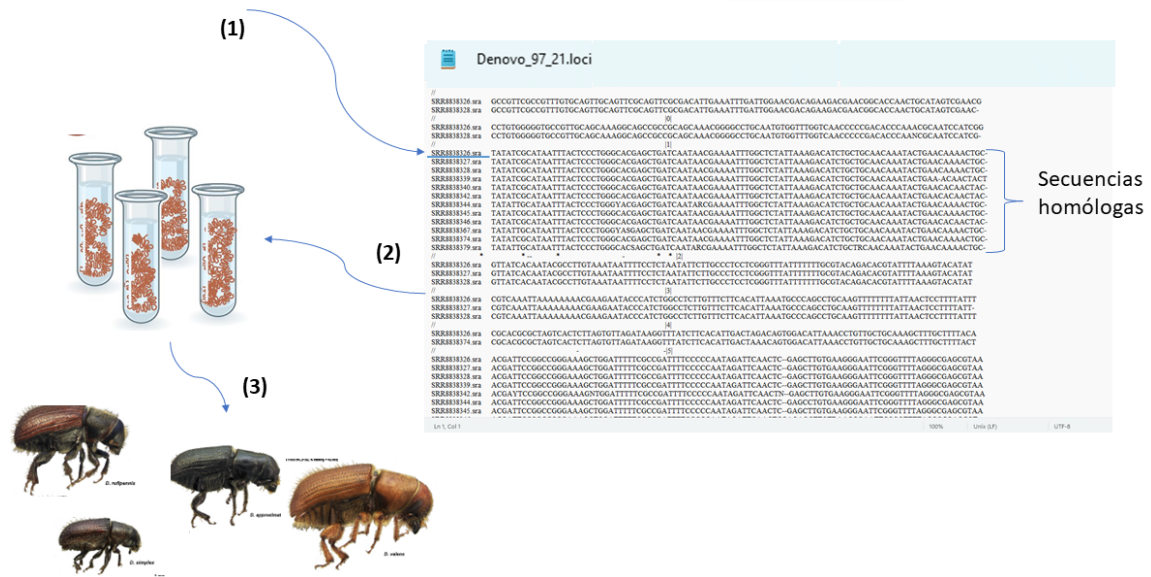
Se presentaron dos tipos de mapas de calor dependiendo la naturaleza de los datos recuperados por cada ensamble y de lo que se quería representar: a) de abundancia relativa, realizados con los paquetes de R *vegan* (Okhasen *et al.*, 2013) y *gplots* (Warnes *et al.*, 2016); b) de abundancia absoluta por medio del paquete *vegan* y *phyloseq* (McMurdie *et al.*, 2013). Ambos tipos de mapas dependen del cálculo de una matriz de distancias para la cual se seleccionó el método de disimilitud de Bray-Curtis debido al papel que tiene la identidad de los taxa comparados en este método, el cual permite utilizar datos con una gran cantidad de observaciones nulas o ceros (Legendre y Legendre, 2012). Los dendrogramas construidos para clasificar a las especies de *Dendroctonus* de acuerdo a las semejanzas entre sus loci microbianos/informativos, se realizaron mediante el método de agrupamiento jerárquico promedio.

Para entender la construcción de los mapas de calor es necesario clarificar que la asignación taxonómica y el blast se realizaron utilizando la primera secuencia asociada a cada loci en el archivo .loci mencionado anteriormente (Fig. 5). Esta secuencia (loci) fue utilizada posteriormente para realizar una búsqueda dentro de su archivo de origen con la finalidad de recuperar todas las secuencias asociadas a ese locus (secuencias homólogas) en particular y la especie de *Dendroctonus* de la cual proviene (Fig. 6).

```

./fasta
>DendroDenovo_97_14
TTGCGTGCCCATAAAAACAAAAGAAAGCGCGGTTTCTCATTTAAATTAATTTTTG...

```



**Figura 6.** Flujo de trabajo del programa empleado para contabilizar las secuencias homólogas de determinados taxa originada en las diferentes especies de *Dendroctonus*. (1) Se ubican las secuencias de cierto taxa con las que se realizó la determinación taxonómica en los archivos FASTA. (2) Se realiza una búsqueda a partir de esta secuencia FASTA en los archivos “.loci” ubicada, al principio de su respectivo loci para contabilizar a todas las muestras que presentan secuencias homólogas. (3) Se suman las abundancias reportadas en cada muestra de la misma especie y a partir de estas se construyen los mapas de calor.

Tras analizar los resultados obtenidos, se realizó una investigación bibliográfica para establecer una posible relación entre los microorganismos encontrados y los hospederos de donde se obtuvo esta información, comparando la diversidad recuperada por medio de los diferentes enfoques presentados.

## RESULTADOS

### Ensamble de datos

A lo largo del ensayo experimental se recuperaron múltiples archivos de texto en diferentes formatos generados como output de los principales procesos bioinformáticos realizados: ensamble de datos de RADSeq, BLAST y determinación taxonómica. A partir de los archivos .stats generados junto con los outputs de ipyrad durante la primera etapa, se registró la cantidad de loci pre y posts filtro. Posteriormente, a partir de los resultados de BLAST se conservaron únicamente aquellas coincidencias con un porcentaje de similitud mayor al 90%, los cuales se categorizaron como “Loci

blast". Para concluir se descartaron los loci pertenecientes a determinados phyla no informativos para reportar a los restantes como la cantidad de loci microbianos informativos. Estos últimos dos pasos se realizaron posteriormente con aquellos loci con más de 97% de solicitud con el fin de realizar la asignación taxonómica a nivel de género (Tabla 2, Fig. 6).

	Ensamble de datos (ipyrad)		BLAST	Asignación taxonómica	
	Loci prefiltro	Loci postfiltro	Loci BLAST	Loci microbianos y de interés taxonómico	% de loci microbianos y de interés recuperados en BLAST
Denovo_97	1,171,307	1,141,059	182,695	9,104	4.983
Denovo_95	1,014,315	988847	16818	569	3.383
DRef_97	587,918	572084	20810	7168	34.445
DRef_95	541,868	527911	19262	7122	36.974
Dpond_97	2,522	2350	279	19	6.810
Dpond_95	2,516	2310	259	16	6.178
GRef_97	213	191	44	5	11.364
GRef_95	194	162	40	4	10.00
G_97_m4	236	211	49	6	12.245
G_97_m2	713	673	347	295	85.014
G_90_m2	786	695	397	352	88.665
G_85_m2	810	699	417	376	90.168

**Tabla 2.** Resultados obtenidos por los diferentes ensambles. La diferencia entre las primeras dos columnas refleja la cantidad de loci retenidos a través de los diferentes pasos del ensamble que se realizan con el pipeline ipyrad. La columna 3 muestra el número de loci recuperados por BLAST (>90% similitud) y la 4 reporta la cantidad correspondiente a loci microbianos. Finalmente la columna 5 muestra el porcentaje de loci que corresponden a loci microbianos y de interés.

El tipo de ensamble más exitoso en términos de la retención de loci a lo largo del pipeline de ipyrad y en etapas posteriores del flujo de trabajo, fueron los de tipo *Denovo* (Figura 7, A). El

ensamble que recuperó la mayor cantidad de loci fue *Denovo\_97* (182,505) seguido de *Denovo\_95* (16,818). También, como se observa con mayor claridad en la tabla 2, estos dos ensambles obtuvieron la mayor cantidad de loci microbianos, particularmente *Denovo\_97* con 9,104 loci.

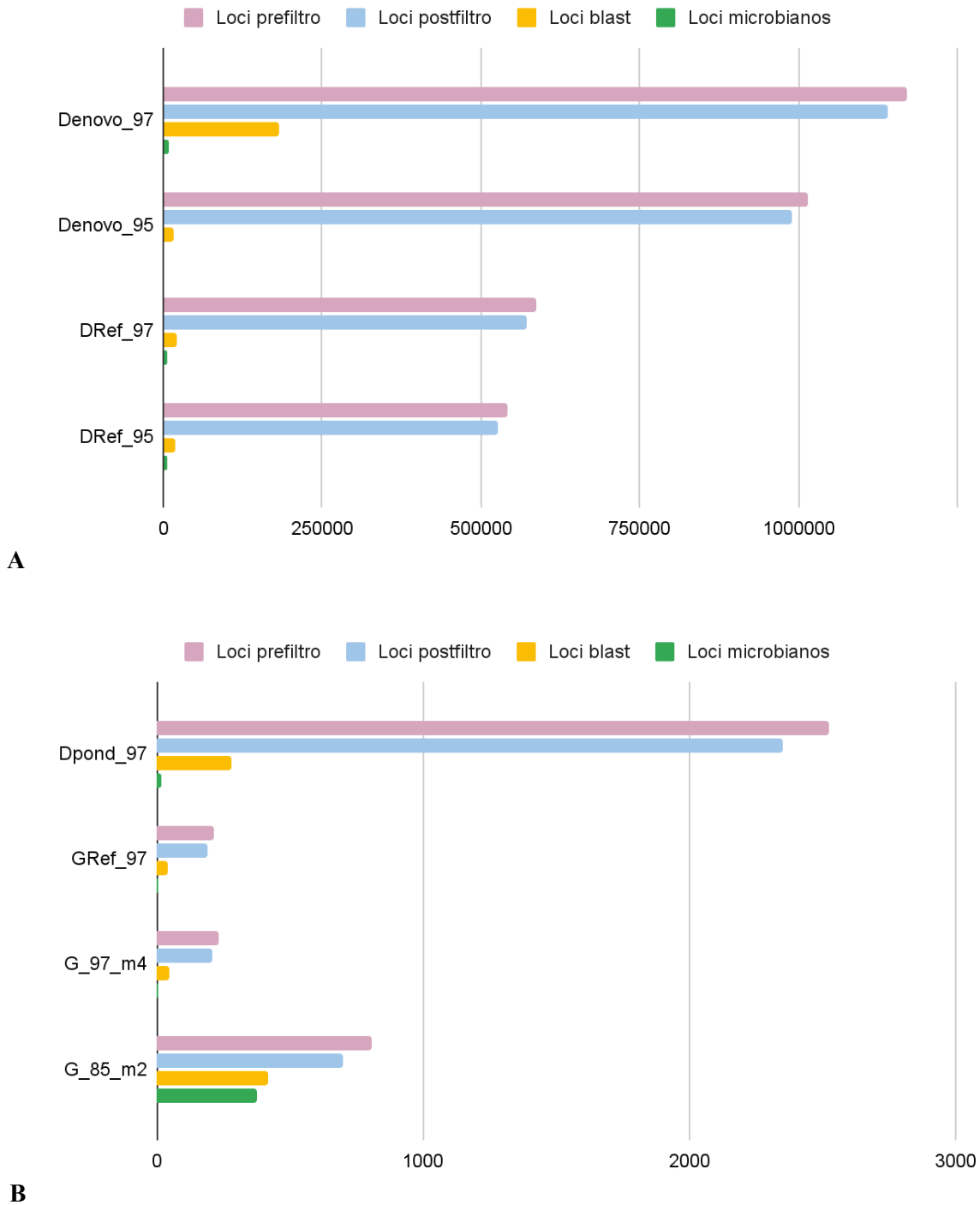
Por otra parte, a pesar de que los ensambles *DRef* también recuperaron una cantidad de loci muy alta, se diferencian del método *Denovo* por recuperar al menos 30% más loci microbianos con respecto a los loci totales recuperados por BLAST, debido a la naturaleza del método, el cual con base en una secuencia de referencia descarta desde pasos tempranos del flujo de trabajo, los loci pertenecientes a *Dendroctonus*.

En el caso de los ensambles realizados usando el método *Denovo - ref*, se evaluó la capacidad de recuperar información al realizarse utilizando únicamente los archivos de RADSeq correspondientes a las cuatro muestras de la especie de referencia *D. ponderosae* (ensambles denominados *DPond*). Bajo el entendido de que mientras más parecidos sean los loci de las muestras a la secuencia de referencia, mejor se descartarían aquellas secuencias correspondientes al genoma del escarabajo, para dar lugar al ensamble de la información de la microbiota.

Los siguientes ensayos recuperaron una cantidad de loci al menos dos órdenes de magnitud menor (Figura 7, B). Los ensambles *DPond*, realizados únicamente con las lecturas correspondientes a cuatro muestras de *D. ponderosae*, recuperaron un número de loci mucho menor en comparación con el ensamble *DRef*. No obstante, como se muestra en la tabla 2, tanto *DPond\_97* como *DPond\_95* presentan una tendencia diferente a la de *DRef*, al recuperar un porcentaje de loci microbianos con respecto a los loci de BLAST bastante menor (6.8% en *DPond\_97* a comparación de 36.9% en *DRef\_97*). Estos resultados pueden servir como guía para priorizar la cantidad de muestras sobre la identidad de estas en investigaciones que recurren a este método para ensamblar datos de diferentes especies.

En cuanto a los ensambles que emplearon una referencia para ensamblar los loci más parecidos a esta, el ensamble *GRef\_97\_m4* con el valor de *mindepth\_majrule* ajustado a 4 no retuvo un número de loci muy diferentes al de los ensambles *+referencia* con la *majority rule* fijada al valor estándar de 6 (GRef\_97: 213, GRef\_95: 194, GRef\_97\_m2: 236). Sin embargo, al disminuir este parámetro a 2, la cantidad de loci microbianos obtenidos aumentó. En la tabla 2 se observa que el ensamble *GRef\_97\_m2* recuperó casi el triple de loci prefiltrado más que los ensambles *GRef* anteriores y que finalmente se identificaron 295 loci microbianos, casi 60 veces más que al realizar el mismo ensamble con la *mindepth\_majrule* fijada en 6. Posteriormente en los ensambles *GRef\_90\_m2* y *GRef\_85\_m2*, esta tendencia a recuperar loci microbianos continuó aumentando, alcanzando hasta los 376 loci microbianos en el ensamble *GRef\_85\_m2* con el valor de CT fijado en 0.85. Estos tres

ensambles, realizados con un valor de *mindepth\_majrule* de 2, presentan el mayor porcentaje de loci microbianos recuperados por BLAST, alcanzando un valor de hasta 90.16% en *GRef\_85\_m2* (Tabla 2).

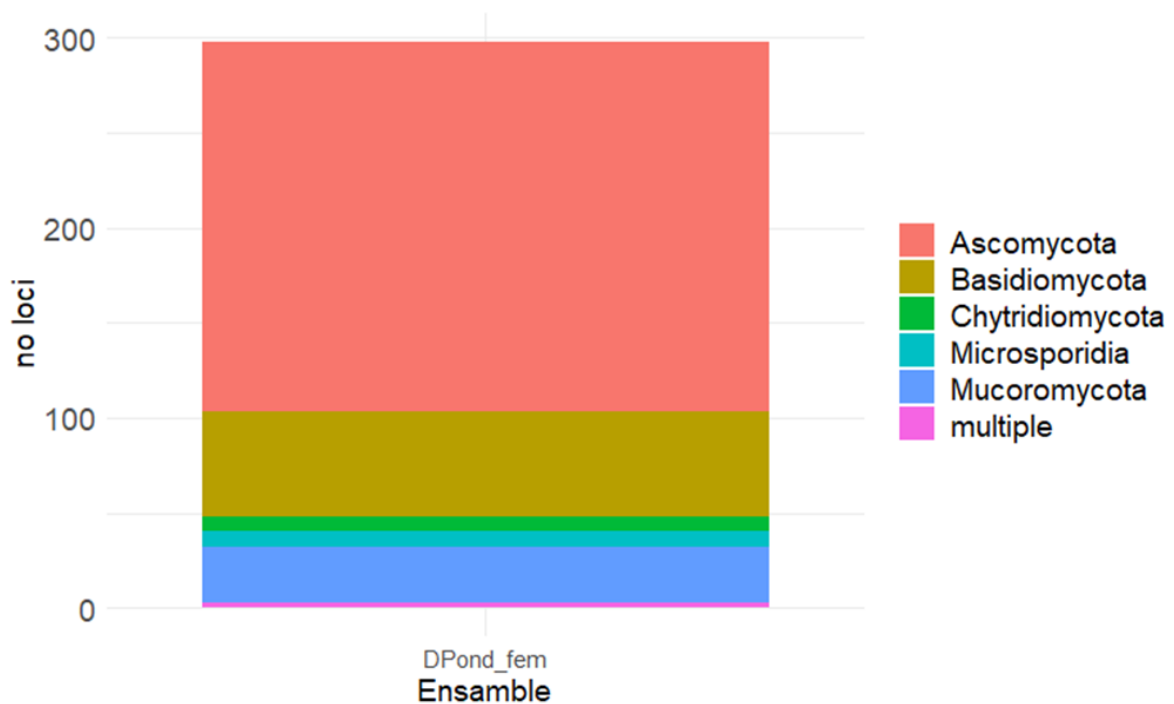


**Figura 7.** Número de loci retenidos a lo largo del ensayo por cada ensamble. La gráfica A incluye a los ensambles que recuperaron la mayor cantidad de loci a lo largo de los diferentes filtros (*Denovo* y *DRef*). La

gráfica B incluye a algunos de los ensambles que debido al método trabajaron con menos loci durante el flujo bioinformático (*DPond*, *GRef* y *GRef\_m*).

Tras emplear la secuencia de referencia de *D. ponderosae* de un organismo hembra en múltiples ensambles bajo diferentes enfoques, se realizó el análisis de BLAST de esta secuencia en contra de una base de datos de hongos, para evaluar la posible influencia de información genómica microbiana presente *a priori*. Los resultados de este análisis recuperaron 298 loci distribuidos entre los diferentes phyla fúngicos conocidos: 195 de Ascomycota, 55 de Basidiomycota, 8 de Chytridiomycota, 29 de Mucoromycota, de Microsporidia y 8 de múltiples phyla (Fig. 8), lo cual equivale a 298 scaffolds (unidad de ensamble) derivados del genoma publicado, con un tamaño reportado de 204 millones de pares de bases.

Loci fúngicos en genoma draft de *D. ponderosae* hembra, 90% de similitud



**Figura 8.** Loci fúngicos recuperados con más del 90% de similitud a partir de la búsqueda de coincidencias taxonómicas de la secuencia de referencia de *D. ponderosae* en contra de la base de datos del NCBI limitada a hongos. Se encontraron 5 de los 6 phyla que conforman al reino Fungi, destacando a Ascomycota como el más abundante.

## Determinación Taxonómica

El ensamble que recuperó la mayor cantidad de loci fue *Denovo\_97* (182,505). Cabe destacar que los ensambles de tipo *Denovo* y *DRef* no sólo obtuvieron mayor número de loci a comparación de los otros dos tipos de ensambles, sino que también recuperaron una mayor diversidad de phyla (Fig. 9).

El phyla más abundante en todos los ensambles fue Arthropoda, sin embargo debido a que estos loci corresponden a información de las diferentes especies de *Dendroctonus*, se filtró este y otros loci como no informativos. De los 67 phyla encontrados en los diferentes ensambles realizados, se descartaron 14 para las comparaciones de diversidad por no pertenecer a grupos relacionados funcionalmente con el contexto biológico de los escarabajos (Brachiopoda, Chordata, Cnidaria, Echinodermata, Hemichordata, Mollusca, Porifera, Platyhelminthes, Priapulida, Rhodophyta) o por no aportar información útil para la clasificación taxonómica (Arthropoda, multiple, no phylum). La categoría “multiple” es asignada en diferentes niveles taxonómicos a aquellos loci que presentan un mismo valor de similitud a diferentes taxa de la base de datos de referencia, esto está relacionado con el tamaño de las secuencias y la resolución de la asignación taxonómica.

El ensamble que recuperó más loci y más categorías de phyla diferentes fue *Denovo\_97*, con 9,104 loci microbianos y de interés. En este ensamble destacaron por su abundancia algunos phyla informativos como Proteobacteria (5,543), Streptophyta (1,998), Nematoda (463) y Apicomplexa (208), además de algunos otros phyla no informativos como Arthropoda (16,941) y Chordata (4,313). También cabe mencionar que entre los ensambles realizados, se encontró un menor grado de diversidad a nivel de phyla en las aproximaciones *De novo - referencia* restringidas a las muestras de *D. ponderosae* y los ensambles + *Referencia* que ocuparon el genoma del hongo simbionte *G. clavigera*.



	<b>Denovo _97</b>	<b>Denovo _95</b>	<b>DRef _97</b>	<b>DRef_ 95</b>	<b>DPond _97</b>	<b>DPond _95</b>	<b>GRef _97</b>	<b>GRef_ 95</b>	<b>GRef_97 _m4</b>	<b>GRef_97_ m2</b>	<b>GRef_90 _m2</b>	<b>GRef_85_ m2</b>
Apicomplexa	208	26	149	147	0	0	0	0	0	0	0	2
Arthropoda	167941	15618	10677	9433	248	232	31	29	36	41	24	22
Ascomycota	170	21	91	92	1	1	3	3	4	292	350	370
Bacteroidetes	107	16	84	77	0	0	0	0	0	0	0	0
Basidiomycota	28	5	13	12	0	0	0	0	0	0	0	0
Chordata	4313	500	2429	2194	9	9	1	1	2	2	5	6
Firmicutes	196	14	142	131	1	1	0	0	0	0	0	0
Mollusca	277	24	118	112	1	1	0	0	0	0	0	0
Nematoda	463	59	257	227	8	7	0	0	0	0	0	0
Proteobacteria	5543	155	4879	5031	2	2	0	0	0	1	1	2
Streptophyta	1998	236	1313	1189	4	4	1	0	1	1	0	0

TOTAL (n = 67 categorías de phyla)	182505	16818	20810	19262	279	259	44	40	49	347	396	417
------------------------------------	--------	-------	-------	-------	-----	-----	----	----	----	-----	-----	-----

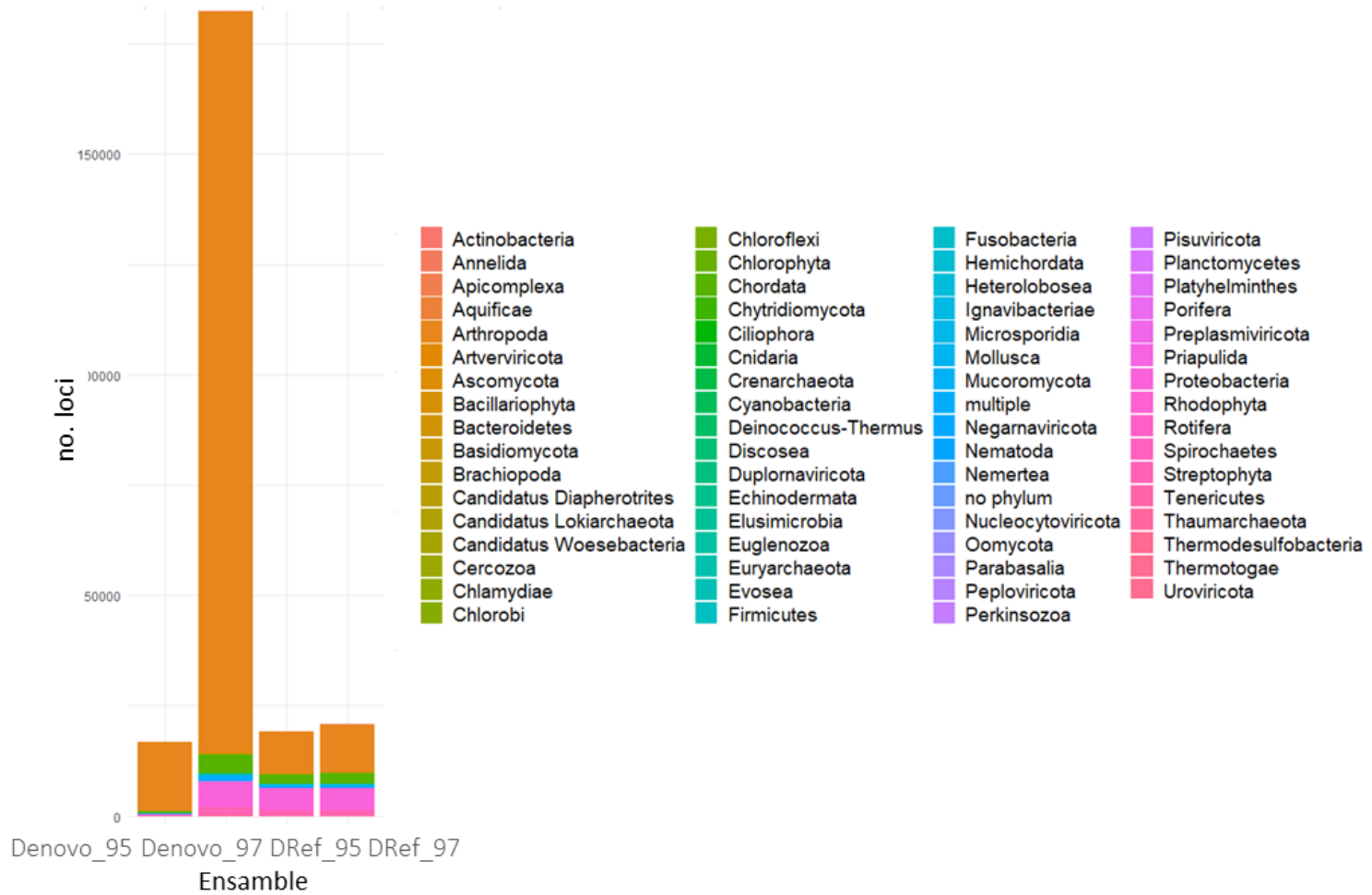
**Tabla 3.** Principales phyla recuperados con un 90% de similitud en la categoría de “loci blast”, previo al filtro de phyla no informativos. La tabla completa con las 67 categorías de phyla diferentes registradas se encuentra en el Anexo 5.

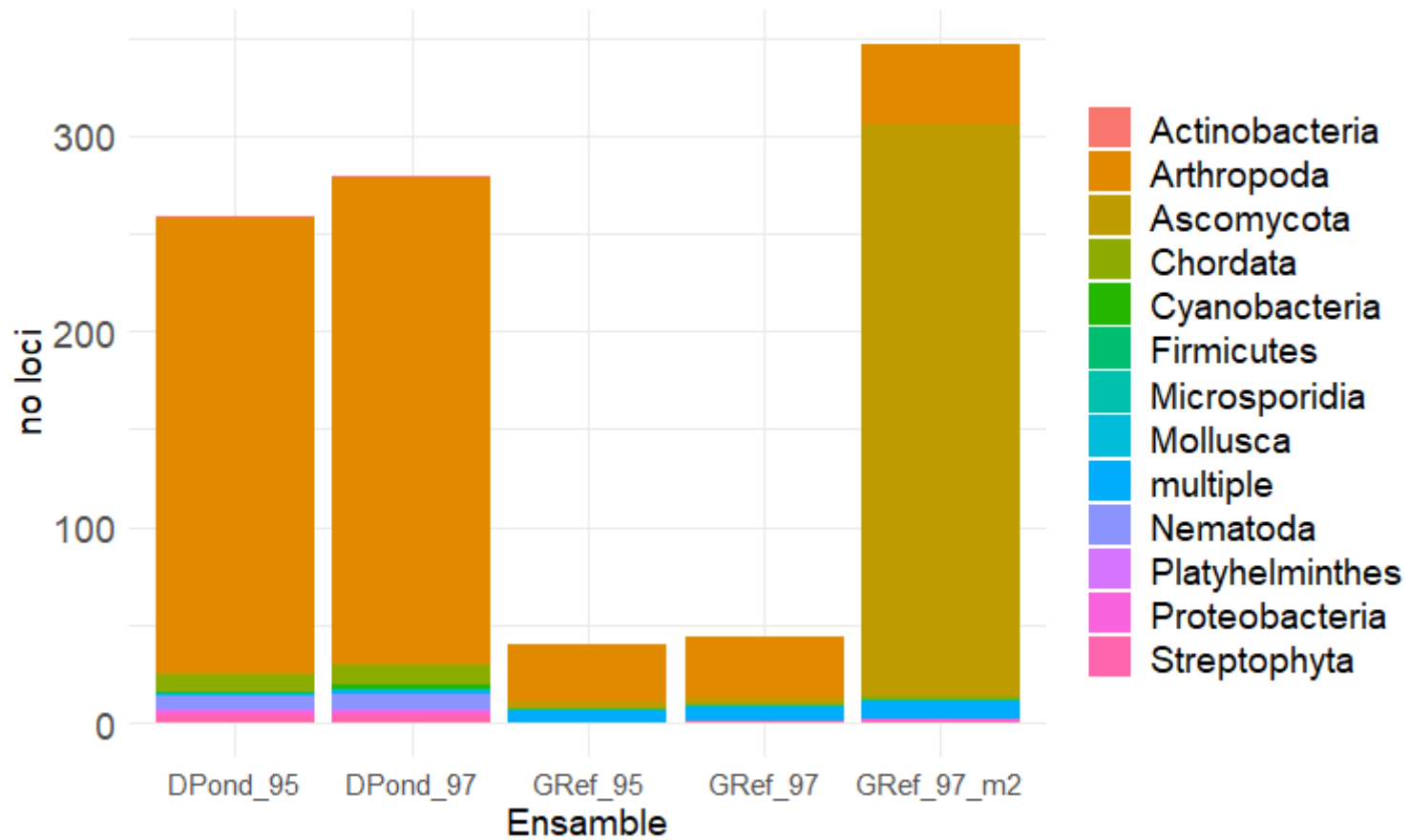
Como puede observarse en la figura 8 y en las Tabla 2 y 3, el ensamble que obtuvo una mayor cantidad de loci con un porcentaje de similitud mayor a 90% fue *Denovo\_97*, alcanzando los 182,505 loci repartidos entre 67 categorías de phyla, lo cual contrasta con aquellos ensambles realizados con un grupo de datos disminuido (*DPond*) y aquellos realizados usando una referencia específica (*GRef*), aunque estos últimos recuperan una mayor cantidad de loci fúngicos. En este sentido, los loci totales recuperados por los ensambles *GRef*, aumentaron dos órdenes de magnitud al disminuir el parámetro *mindepth\_majrule* a 2.

En cuanto a la evaluación del valor de CT, y su influencia en la recuperación de un mayor número de loci, en la mayoría de los ensambles, exceptuando aquellos con la *mindepth\_majrule* fijada en 2, los valores del parámetro fueron más eficientes en el extremo superior del rango utilizado (97% de similitud) (Fig. 9).

Cabe resaltar que el ensamble *DRef\_95* recuperó menos loci totales que *Denovo\_95*, debido a que se filtra la información de *Dendroctonus* como parte del proceso de ensamble en *ipyrad*, sin embargo recupera más loci de los otros phyla considerados potenciales integrantes de su microbiota. Esto no sucede en el caso de los ensambles *DRef* y *Denovo* con CT de 97, sugiriendo un mejor rendimiento en el caso de este método de ensamble con un valor de 95% de similitud.

### Loci BLAST totales, 90% de similitud

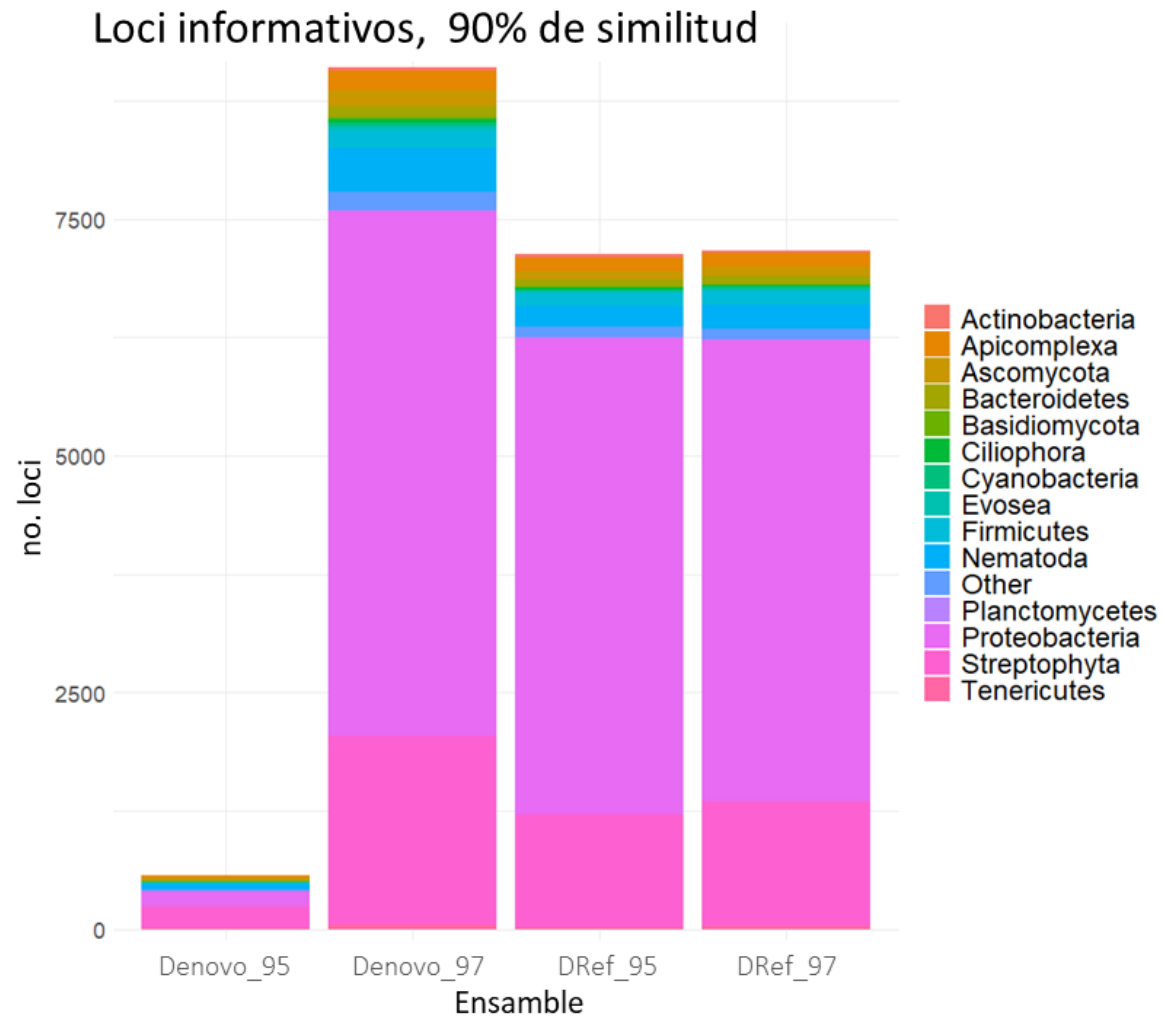




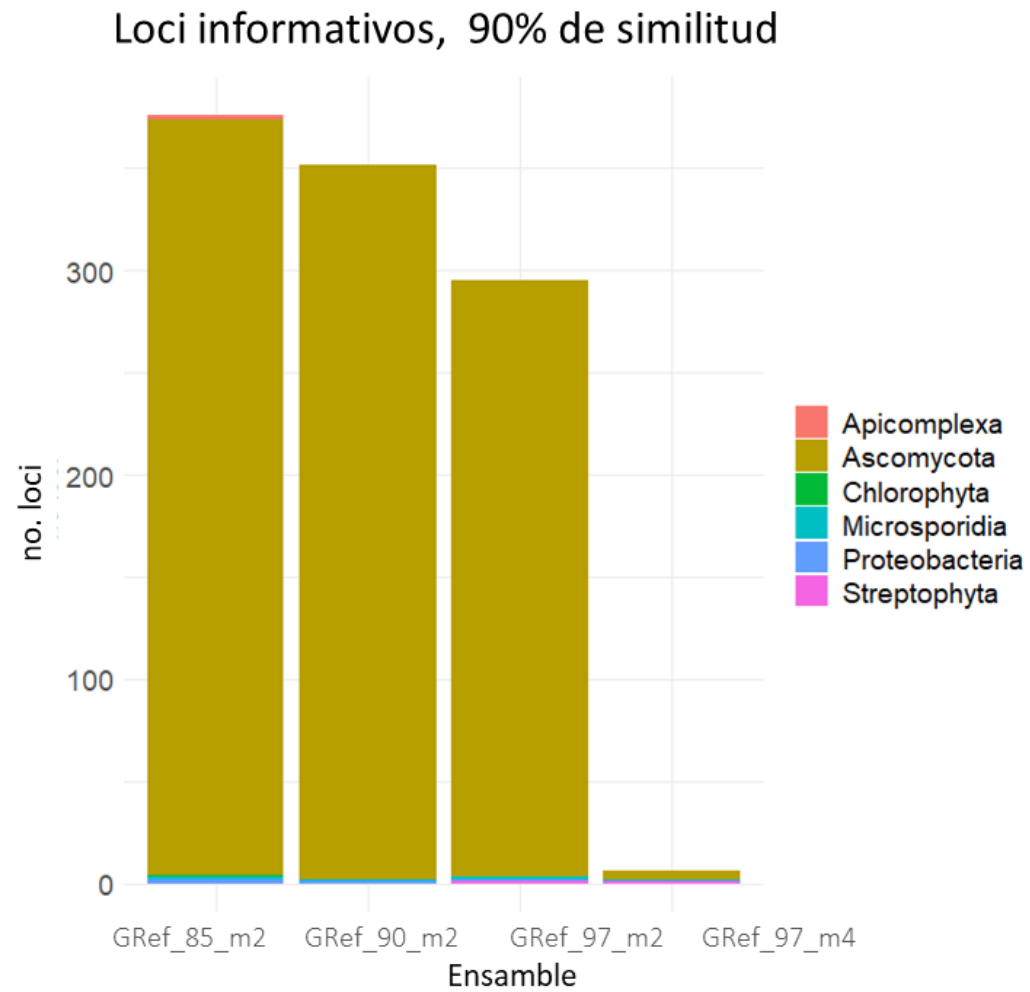
**Figura 9.** Total de phyla encontrados en cada tipo de ensamble previo a la eliminación de phyla no informativos. La gráfica A incluye al grupo de ensambles que recuperó la mayor cantidad de loci y hasta 67 categorías de phyla (*Denovo* y *DRef*). La gráfica B incluye a representantes de los ensambles que recuperaron menos loci y sólo 13 categorías de phyla (*DPond*, *GRef*, *GRef\_m*).

A partir de la depuración de aquellos phyla que no mantienen una relación ecológica con el sistema, se reportaron un total de 53 phyla informativos por medio del ensamble más exitoso (*Denovo\_97*). Con base en este ensamble se recuperaron 5,543 loci de Proteobacteria, posicionando a este grupo como el phylum más abundante, seguido de Streptophyta con 1,998 y de Nematoda con 463. Entre los otros phyla se destacan aquellos pertenecientes al reino Fungi por la variedad de relaciones ecológica que mantienen con *Dendroctonus*: Ascomycota recuperó 170 loci, Basidiomycota 28, Chytridiomycota 2, Mucoromycota 8 y Microsporidia 8 (Fig. 10).

Por su parte, los ensambles *GRef* destacaron desde etapas tempranas de este trabajo, ya que fueron los primeros y únicos en presentar phyla fúngicos (Ascomycota) como grupos dominantes ante el resto de la diversidad recuperada. Debido a esta particularidad y al objetivo de estudiar la relación simbiótica obligada característica de múltiples especies de este género de escarabajo con hongos micangiales, se evaluó el efecto que tenía un cambio en el parámetro *mindepth majority rule* en este ensamble y su influencia en la cantidad de loci informativos resultantes (Fig. 11). A partir de esto, observamos el mismo comportamiento en todos los ensambles *GRef\_m2* con respecto a la dominancia de una categoría (Fig. 11). El ensamble más exitoso, *GRef\_85\_m2*, concentró un 98% de loci de Ascomycota (370 loci), con el 2% restante repartido entre Apicomplexa, Chlorophyta, Microsporidia y Proteobacteria (19 loci). De igual forma, la figura 11 ilustra cómo el parámetro *mindepth\_majrule* no tiene una relación proporcional con los resultados obtenidos al recuperar sólo 6 loci microbianos en el ensamble *Gref\_97\_m4*, en comparación con el aumento brusco de loci recuperados al disminuir este valor a 2. En contraste con esto, el CT parece mantener una relación inversamente proporcional con la cantidad de loci informativos en los ensambles *GRef\_m2*, bajo el intervalo evaluado.



**Figura 10.** Phyla informativos por tipo de ensamble. Tras filtrar los loci de catorce categorías de phyla, los ensambles de la gráfica A retuvieron 53 categorías de phyla, de las cuales 40 fueron agrupadas en la categoría de “Other” por su baja abundancia, mientras que en la gráfica B se recuperaron sólo 8 phyla.



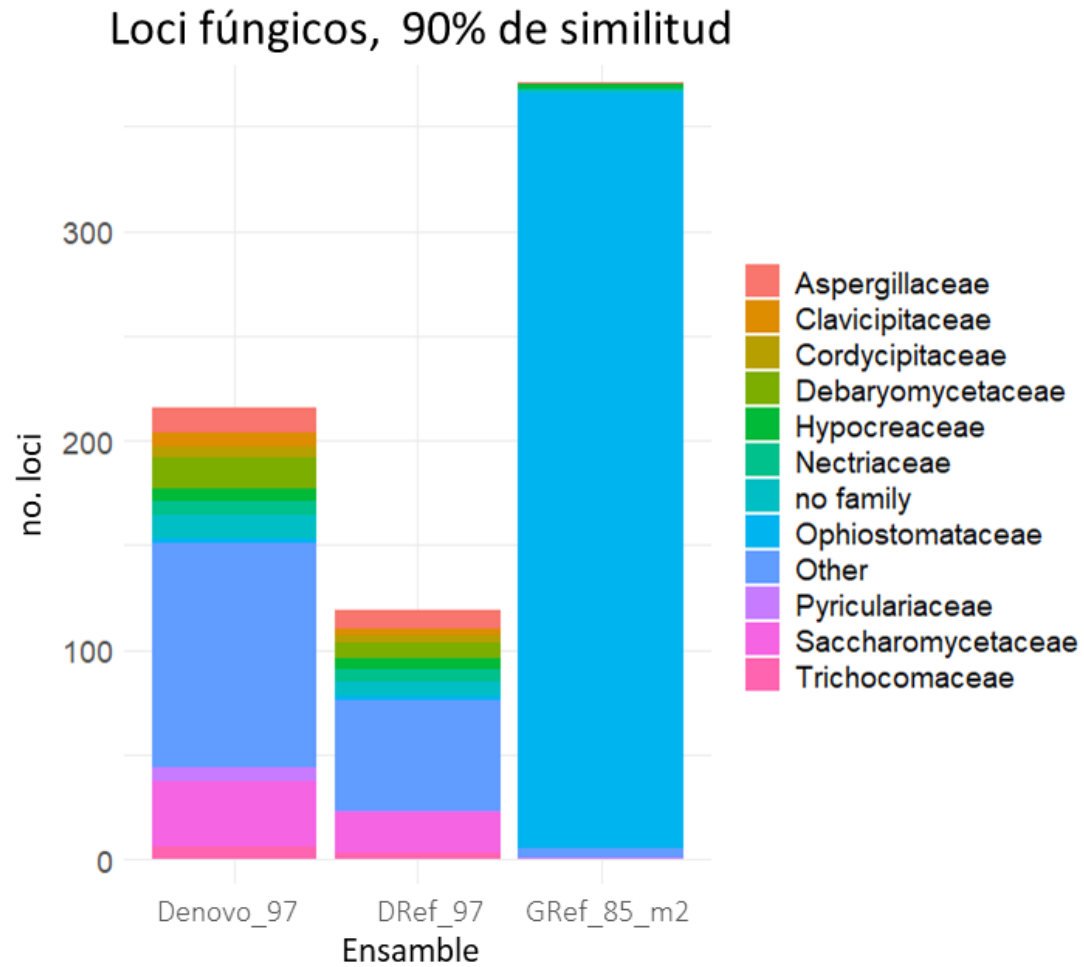
**Figura 11.** Phyla encontrados en los ensambles *GRef* con el parámetro *mindepth\_majrule* modificado. Los ensambles con un valor de m2 recuperaron aproximadamente 300 loci más que cuando se modifica este parámetro a 4 o cuando se mantiene en el valor estándar de 6.

## Taxa de interés

Tras explorar los loci informativos recuperados por los diferentes ensambles, se seleccionaron los ensambles más exitosos en términos del número de loci recuperados, para estudiar la diversidad de familias recuperada en algunos grupos de interés por su relación ecológica con los escarabajos del género *Dendroctonus*. Al caracterizar la diversidad del reino Fungi se incluyeron los phyla Ascomycota, Basidiomycota, Chytridiomycota, Glomeromycota, Mucoromycota y Microsporidia. El ensamble que recuperó más loci fúngicos fue *GRef\_85\_m2* con 371 loci, seguido de los 216 loci reportados en *Denovo\_97*, 119 en *DRef\_97*. Sin embargo, el ensamble que recuperó más loci (*GRef\_85\_m2*) no fue el que encontró un mayor diversidad supraespecífica, al presentar únicamente 10 familias (Fig. 16). Entre estas destaca Ophiostomataceae como la familia dominante, debido a que concentra el 97% de los loci encontrados por este ensamble. La mayor parte de los loci pertenecientes a esta familia (~96%) hicieron match contra la referencia del genoma completo de *Grosmania clavigera* empleada para realizar el ensamble + referencia en ipyrad.

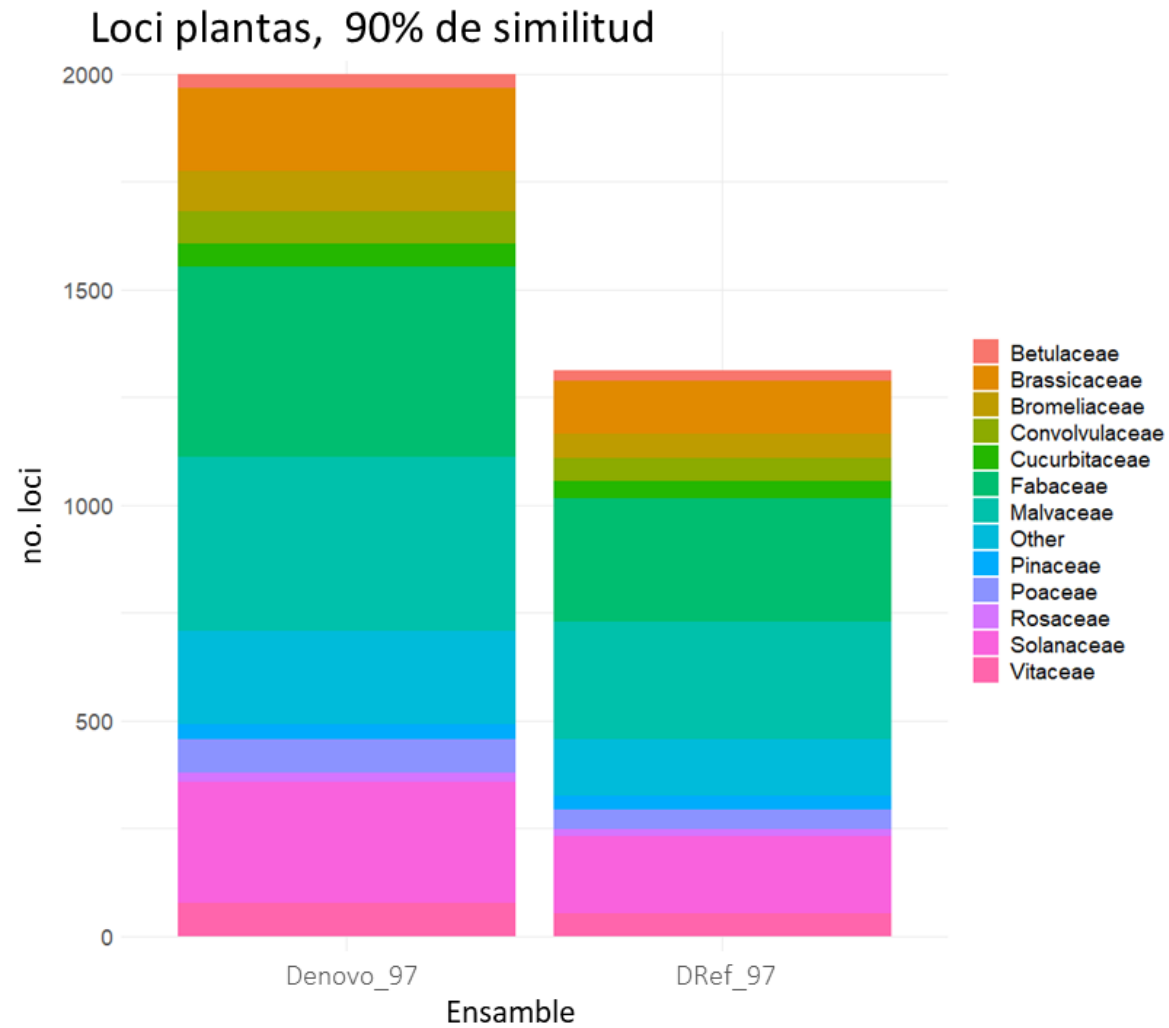
El ensamble que recuperó mayor diversidad de loci de distintos taxa fue *Denovo\_97*, en el cual se contabilizaron 67 categorías de familias de hongos, de las cuáles la mayor parte se agruparon en la categoría de “Other” por su baja abundancia. Tanto en este ensamble como en *DRef\_97*, la familia más abundante de estas fue Saccharomycetaceae, reportando 31 y 19 loci respectivamente, seguida de Debaryomycetaceae y Aspergillaceae (Fig. 12).





**Figura 12.** Familias de hongos recuperadas por los ensambles más exitosos. Se encontraron 67 categorías de familias diferentes concentradas en los ensambles *Denovo\_97* y *DRef\_97*, de las cuales 54 fueron agrupadas en la categoría de “Other” por su baja abundancia. El ensamble *GRef\_85\_m2* registró únicamente 10 categorías de familia.

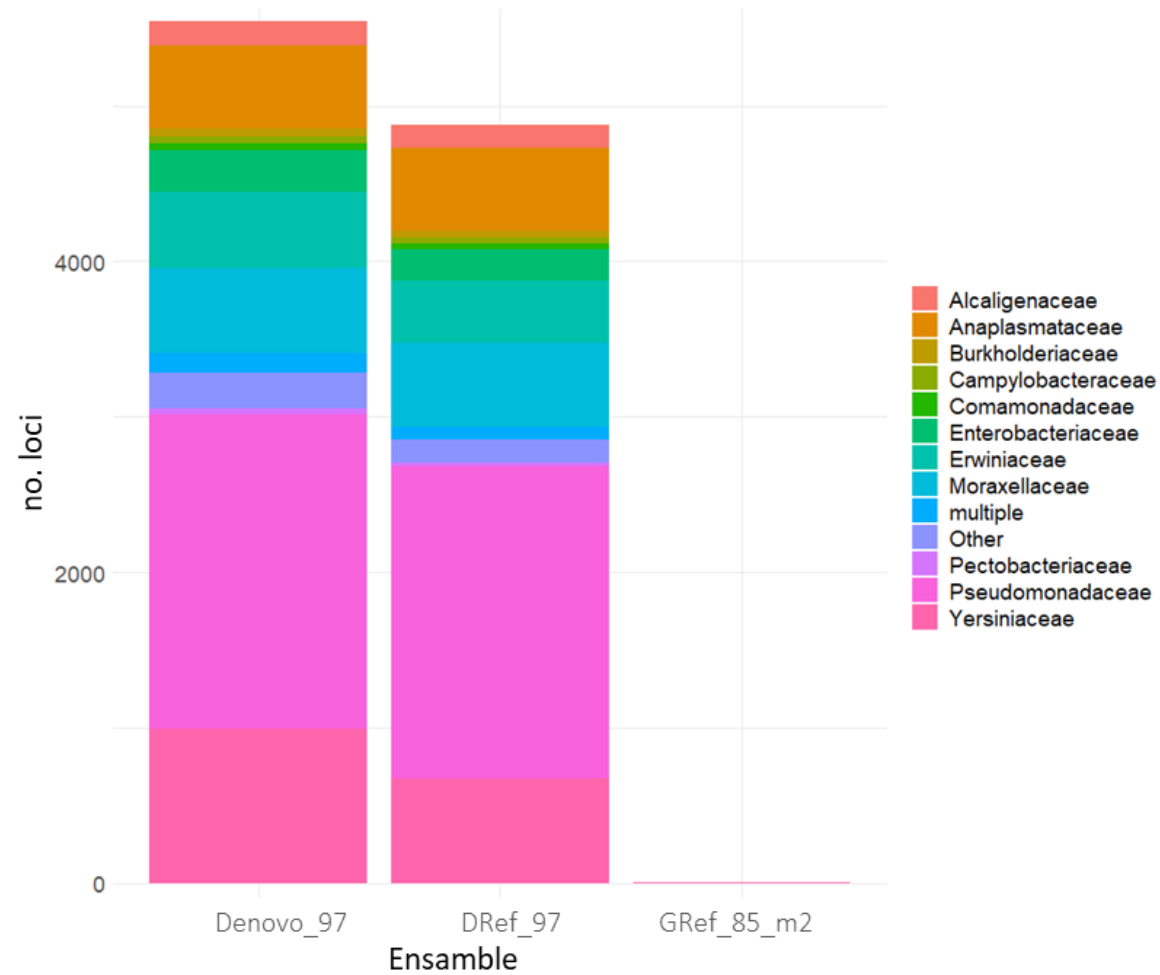
Dentro del phylum Streptophyta (plantas terrestres y algas carofitas), el ensamble *Denovo\_97* permitió recuperar 1,998 loci y 74 familias, mientras que *DRef\_97* recuperó 1,313 loci y 63 familias. La familia más abundante en ambos ensambles fue Fabaceae con 442 loci registrados en el ensamble más exitoso, seguida de Malvaceae con 403 y Solanaceae con 280. Entre las familias consideradas como informativas por su relación biológica con el género *Dendroctonus* se encuentra Pinaceae, de la cual se reportó una abundancia de sólo 34 loci. En el ensamble *GRef\_85\_m2* no se encontraron loci pertenecientes al phylum Streptophyta (Fig. 13).



**Figura 13.** Familias de plantas recuperadas por los ensambles más destacados. Se encontraron 74 categorías de familias concentradas en *Denovo\_97* y *DRef\_97*, de las cuales 61 fueron agrupadas en la categoría de “Other” por su baja abundancia.

Al analizar los resultados de las familias recuperadas de Proteobacteria, se obtuvieron mejores resultados por medio del ensamble *Denovo\_97*, en el cual se encontraron 5,543 loci distribuidos en 67 categorías de familia, de las cuáles las más abundantes fueron Pseudomonadaceae con 2,030 loci, Yersinaceae con 989 y Moraxellaceae con 547. El ensamble *DRef\_97* obtuvo una distribución de loci muy similar entre las familias dominantes; recuperó 4,879 loci y 61 familias en total. Entre algunas otras familias destacadas se encuentra Anaplasmataceae con 533 loci y Enterobacteriaceae con 257. El ensamble *GRef\_95\_m2* recuperó únicamente 2 loci pertenecientes a dos familias de proteobacterias diferentes (Fig. 14).

### Loci proteobacterias, 90% de similitud



**Figura 14.** Familias de proteobacterias recuperadas por los ensambles más destacados. Se encontraron un total de 67 categorías de familias concentradas en los ensambles *Denovo\_97* y *DRef\_97*, de las cuales 54 fueron agrupadas en la categoría de “Other” por su baja abundancia. El ensamble *GRef\_85\_m2* registró sólo 2 categorías de familia.

## Tamaño de los loci

Debido a la influencia del tamaño de las secuencias utilizadas para el análisis de BLAST, se recuperaron las medidas de tendencia central y desviación estándar del tamaño de los loci utilizados para caracterizar taxonómicamente a nivel del phyla en el caso de los ensamblajes *Denovo\_97* y *Grf\_85\_m2*. El primero debido a la alta diversidad de phyla recuperados, y el segundo por la importancia biológica de la simbiosis entre algunos grupos de hongos y los escarabajos del género *Dendroctonus* (Tabla 4). Se realizó la comparación del tamaño de los loci de los phyla compartidos entre estos dos ensamblajes y se incluyeron algunos phyla no informativos en el caso del ensamblaje *Denovo\_97*, con la finalidad de explorar la relación entre la asignación taxonómica y el tamaño de los loci.

A partir de las medidas calculadas es posible identificar una aparente tendencia hacia un tamaño y un rango de desviación estándar. Esta se ve reforzada al notar que aquellos phyla que presentan valores de tamaños muy pequeños recuperados por *GRef\_85\_m2* (Apicomplexa y Chlorophyta) son también phyla muy poco representados, con muestras de dos o incluso una secuencia.

De igual manera, se observa una diferencia entre los rangos de los tamaños encontrados por ambos ensamblajes. Mientras que *GRef\_85\_m2* registró un tamaño mínimo de 28 en Apicomplexa y Chlorophyta, y un máximo de 146 en Arthropoda, en el ensamblaje *Denovo\_97* el tamaño mínimo fue de 75 y se encontró en la categoría de *multiple* mientras que el máximo fue de 126 proveniente de Arthropoda (Fig 15).

En ambos ensamblajes se observa que las barras de desviación estándar no se superponen entre ellas, manteniéndose en un rango de entre 0.341 a 3.19 en *Denovo\_97* y de 1.882 a 26.33 en *GRef\_85\_m2* (Fig 15). Estos resultados refuerzan la hipótesis sobre la homogeneidad del tamaño de los loci en los diferentes taxa para ambos ensamblajes, es decir, que el tamaño de los loci no parece mantener una relación con el número de loci encontrado en cada taxa, a excepción de dos de las categorías menos abundantes de *GRef\_85\_m2*, Apicomplexa y Chlorophyta.

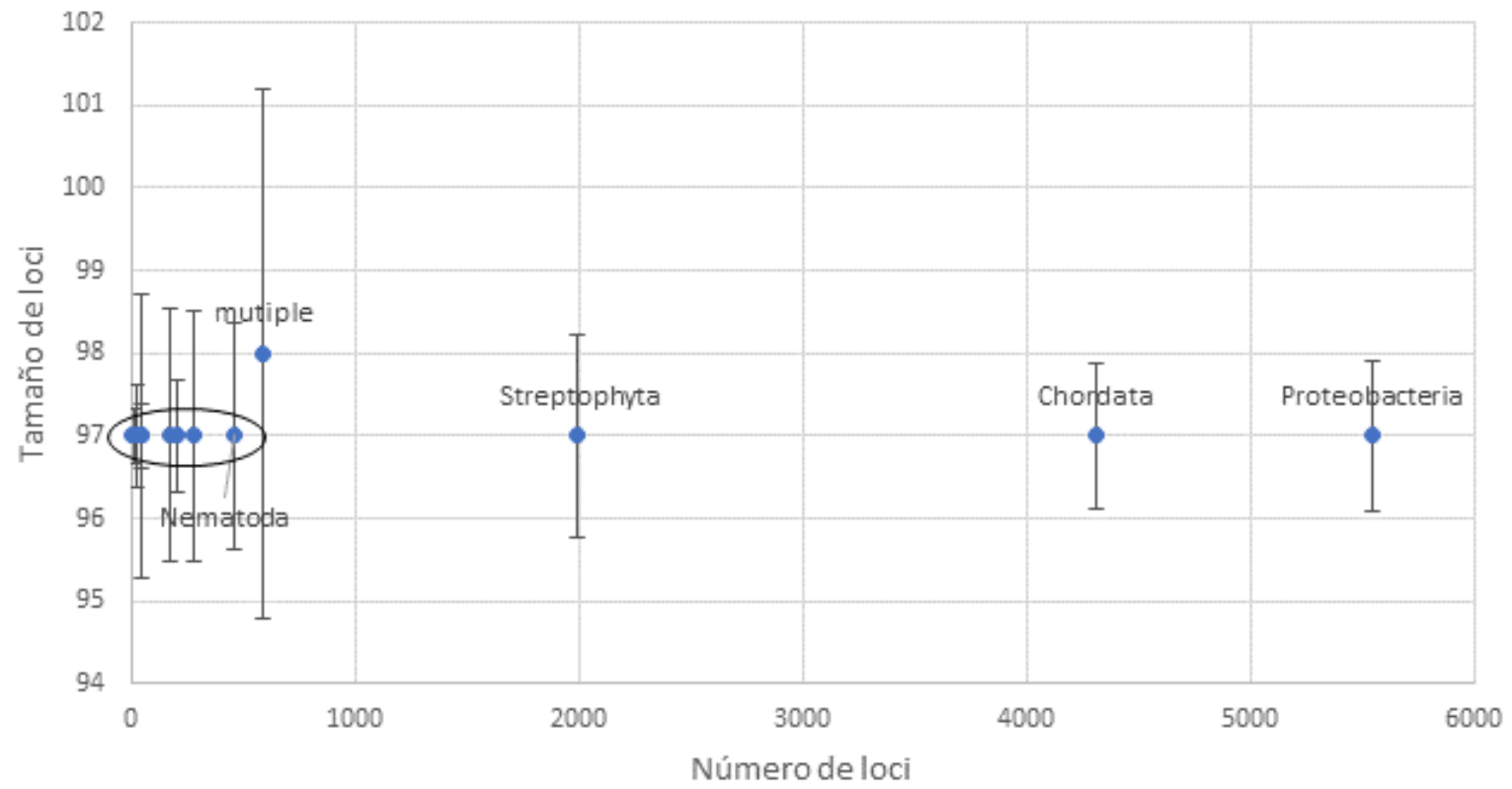
La categoría con un mayor valor de desviación estándar en ambos ensamblajes es “*multiple*”. Cabe resaltar que los valores elevados de la desviación estándar en el ensamblaje *GRef\_85\_m2* se explican a partir de la manera en la que se calcula este valor; como hay una mayor variedad de tamaños de loci sondeados, el resultado final es un mayor valor en el estadístico.

	Denovo_97						GRef_85_m2					
	no. de loci	Media	Mediana	Moda	Rango	D.E	no. de loci	Media	Mediana	Moda	Rango	D.E
Arthropoda	167941	97	97	97	81-126	1.322	22	100	97	97	29 - 146	20.478
Proteobacteria	5543	97	97	97	85 - 121	0.917	2	97	97	97	97	0
Chordata	4313	97	97	97	91 - 121	0.890	6	100	97	97	97 - 113	6.531
Streptophyta	1998	97	97	97	77 - 121	1.234						
multiple	585	98	97	97	75 - 121	3.199	11	97	97	97	27 - 135	26.333
Nematoda	463	97	97	97	82 - 118	1.361						
Mollusca	277	97	97	97	96-120	1.522						
Apicomplexa	208	97	97	97	96 - 106	0.674	2	28	28	28	28	0
Ascomycota	170	97	97	97	95 - 108	1.526	370	97	97	97	96 - 121	1.882
no phylum	51	97	97	97	85 - 99	1.717	2	97	97	97	97	0
Cnidaria	45	97	97	97	97 - 99	0.382						
Basidiomycota	28	97	97	97	97 - 100	0.629						
Oomycota	17	97	97	97	97	0						
Chlorophyta	16	97	97	97	97 - 98	0.341	1	28	28	28	28	
Microsporidia	8	97	97	97	97	0	1	97	97	97	97	

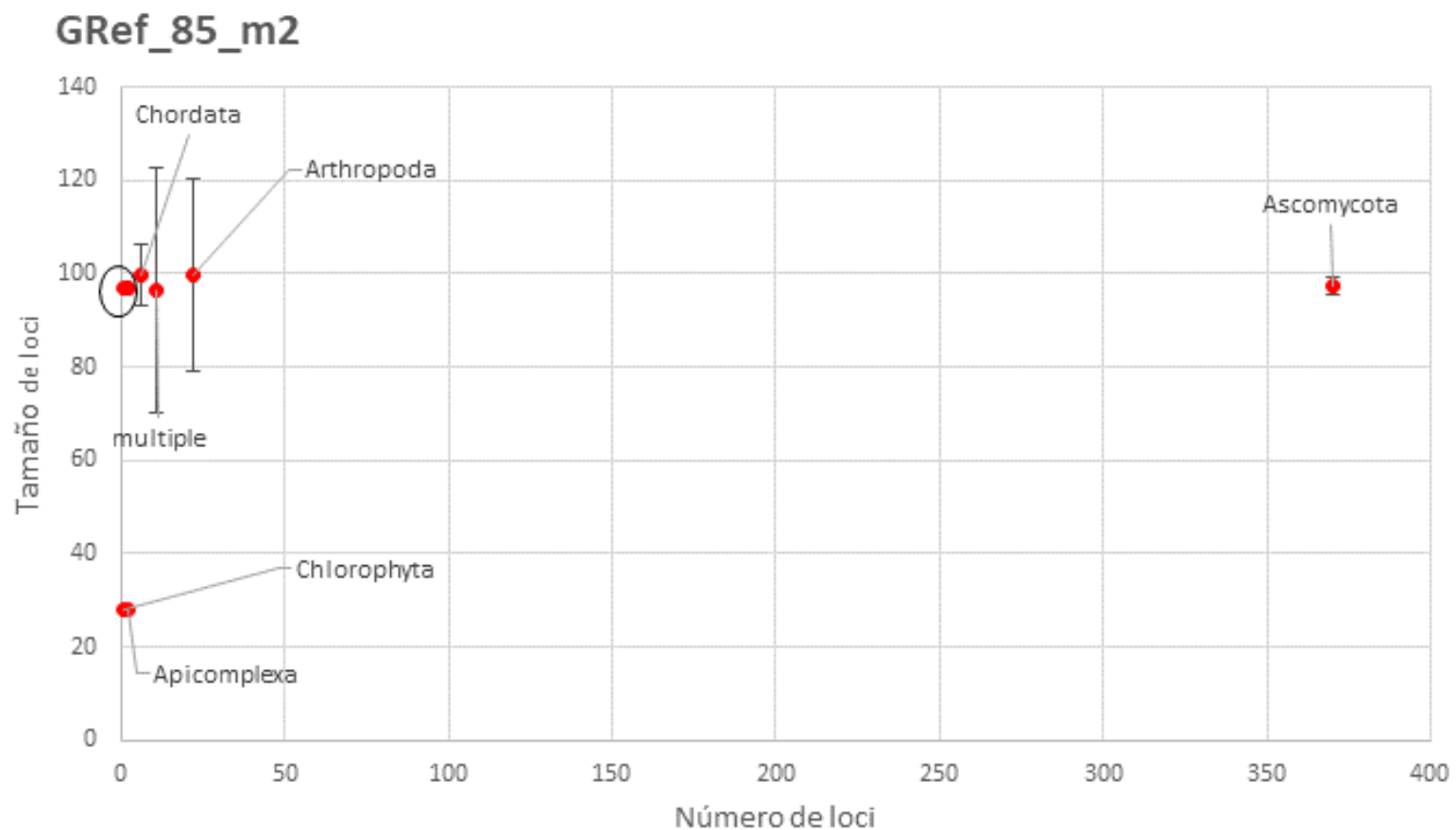
**Tabla 4.** Medidas de tendencia central de los tamaños de las secuencias FASTA asignadas a los diferentes phyla recuperados por los ensambles *Denovo\_97* y *GRef\_85\_m2*.

En ambos grupos de datos se observa un valor único distribuido entre la mayoría de los phyla. La mayor parte de las secuencias de *Denovo\_97* tienen un tamaño de 97 pb reflejado en las media, moda y mediana, mientras que los phyla recuperados por *GRef\_85\_m2* presentaron también este valor con una distribución menos homogénea.

## Denovo\_97







**Figura 15.** Tamaño promedio de los loci con los que se realizó la asignación taxonómica en el ensamble *Denovo\_97* y *GRef\_85\_m2*. Las barras de error asociadas a los diferentes phyla representan la desviación estándar (Tabla 4). En ambas gráficas se observa que el tamaño promedio compartido entre los diferentes phyla es cercano a 97 y que la categoría con mayor desviación estándar es *multiple*. Debido a la diferencia entre el número de loci en la gráfica de *Denovo\_97* no se incluyó a “Arthropoda”. Los

phyla con loci de 97pb encerrados en un círculo negro en el gráfico *Denovo\_97* incluyen a Mollusca, Apicomplexa, Ascomycota, no phylum, Cnidaria, Basidiomycota, Chlorophyta, Oomycota y Microsporidia. Los phyla encerrados en un círculo negro en el gráfico *GRef\_85\_m2* incluyen a no phylum y a Microsporidia.

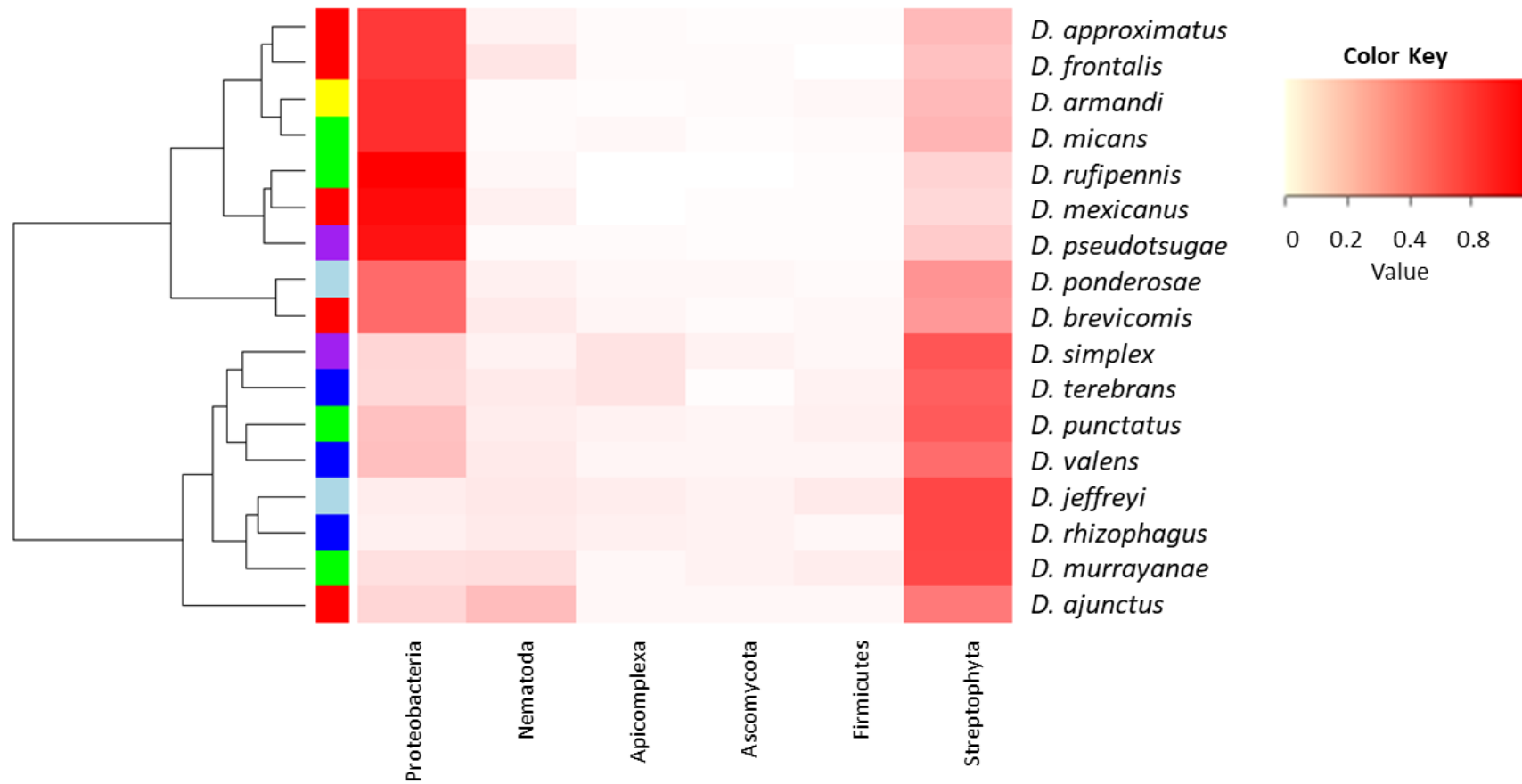
### **Distribución de loci entre las spp. de *Dendroctonus***

Con el fin de visualizar y comparar la abundancia relativa de los loci pertenecientes a los principales phyla informativos por cada especie de *Dendroctonus*, se construyó un mapa de calor y un dendrograma con base en estos valores (Fig. 16) y al método explicado en la figura 6, en el cual se describe la diferencia entre los loci empleados para la asignación taxonómica, y las secuencias homólogas a estos loci provenientes de diferentes muestras correspondientes a las especies de *Dendroctonus*.

Se encontró una distribución heterogénea de las secuencias homólogas entre los 9,104 loci pertenecientes a los phyla informativos recuperados por *Denovo\_97*. A lo largo de *Dendroctonus*, el phylum más numeroso fue Proteobacteria sumando un total de 12,602 secuencias distribuidas entre las diferentes especies, seguido de Streptophyta con 6,827, Nematoda con 1,470, Apicomplexa con 742, Firmicutes con 660 y Ascomycota con 606 (Fig. 16).

Las especies en las que se encontró una mayor cantidad de loci fueron *D. micans*, seguida de *D. mexicanus* y *D. armandi*. Es posible identificar dos pares de clusters pequeños formados por *D. approximatus* y *D. frontalis*, especies de micangio protorácico pertenecientes filogenéticamente al grupo VI y *D. micans* y *D. rufipennis*, especies del grupo filogenético III, posiblemente por la abundancia de secuencias homólogas de Proteobacteria. Más allá de estos resultados, no es posible observar un patrón en la distribución de la diversidad microbiana y los grupos filogenéticos a partir del dendrograma construido.

### Distribución de secuencias homólogas de los principales phyla de *Denovo\_97*



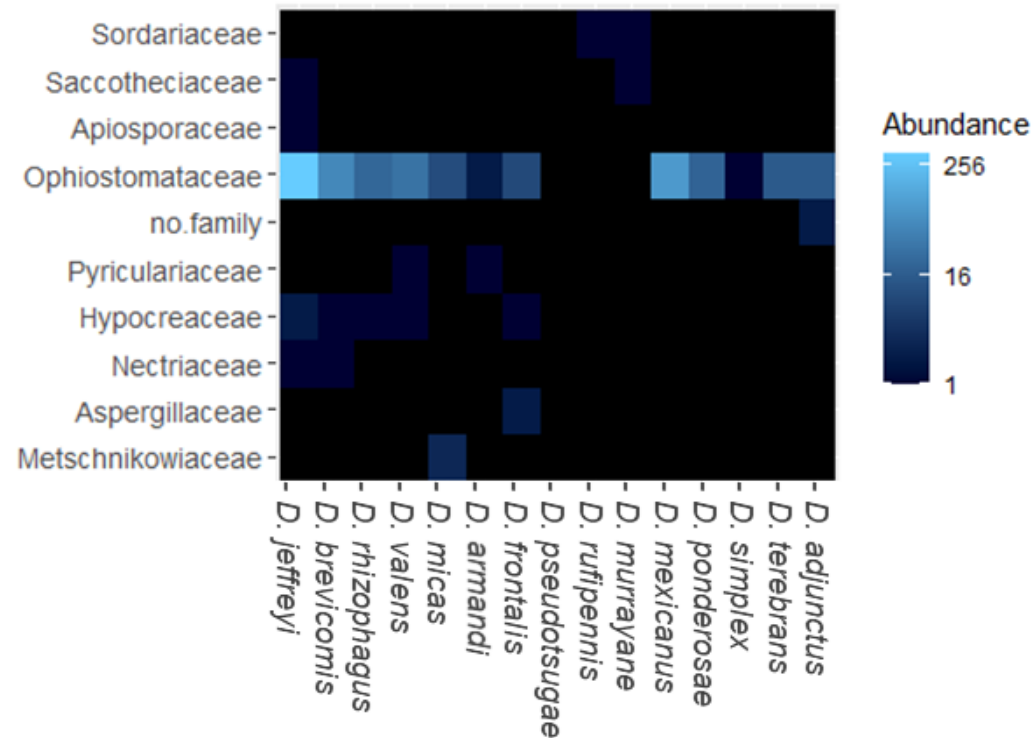
**Figura 16.** Mapa de calor de los principales phyla informativos recuperados en cada especie de *Dendroctonus* a partir del ensamble *Denovo\_97*. El gradiente de color de cada phyla corresponde a su abundancia relativa en los diferentes escarabajos hospederos. Se descartaron a aquellos phyla no informativos y aquellos que no presentaban una

abundancia relativa mayor al 1% en al menos una especie hospedera. Se construyó un dendrograma basándose en los patrones de asociación de los phyla informativos entre las especies de *Dendroctonus* mediante el método de agrupación jerárquica de Bray-Curtis. Estas especies se etiquetaron con los colores asociados a los grupos filogenéticos de la Fig. 6: I Amarillo, II Morado, III Verde, IV Azul claro, V Rojo y VI Azul.

A partir de los 376 loci informativos obtenidos por el ensamble *GRef\_85\_m2*, de los cuales 371 pertenecían al reino Fungi, se buscó presentar la distribución de las secuencias homólogas a nivel de familia (Fig 16). Entre estas se encontró sólo un loci perteneciente a Microsporidia en las muestras de *D. adjunctus*, sin embargo esta no logró ser clasificada hasta nivel de familia por lo que se ilustró en el mapa de calor bajo la categoría “no family”. De las 9 familias de Ascomycota recuperadas, la más abundante fue la familia de simbiontes Ophiostomataceae, de la cual se encontraron 621 secuencias homólogas repartidas entre todas las especies del género a excepción de *D. pseudotsugae* y *D. murrayanae*. Las especies que concentran una mayor cantidad de secuencias homólogas de esta familia fueron *D. jeffreyi* (349), *D. mexicanus* (91) y *D. brevicomis* (56).

El resto de las familias fúngicas encontradas por este ensamble presentan una abundancia absoluta muy baja, la cual no excede a las 3 secuencias homólogas en cada especie, en el mejor de los casos. Al explorar el resto de los datos se encontró que las familias de los otros phyla también se encontraban pobremente representadas a lo largo del género.

## Distribución de secuencias homólogas de las principales familias de Ascomycota de *GRef\_85\_m2*



**Figura 16.** Mapa de calor de las diferentes familias fúngicas encontradas en cada especie de *Dendroctonus* a partir de el ensamble *GRef\_85\_m2*. Este mapa se realizó con el fin de visualizar las abundancias absolutas de cada familia en las diferentes especies hospederas debido a que la mayor cantidad de loci recuperados corresponde a la familia Ophiostomataceae. *D. approximatus* y *D. punctatus* fueron descartadas de este análisis al no reportarse ningún loci proveniente de estas especies.

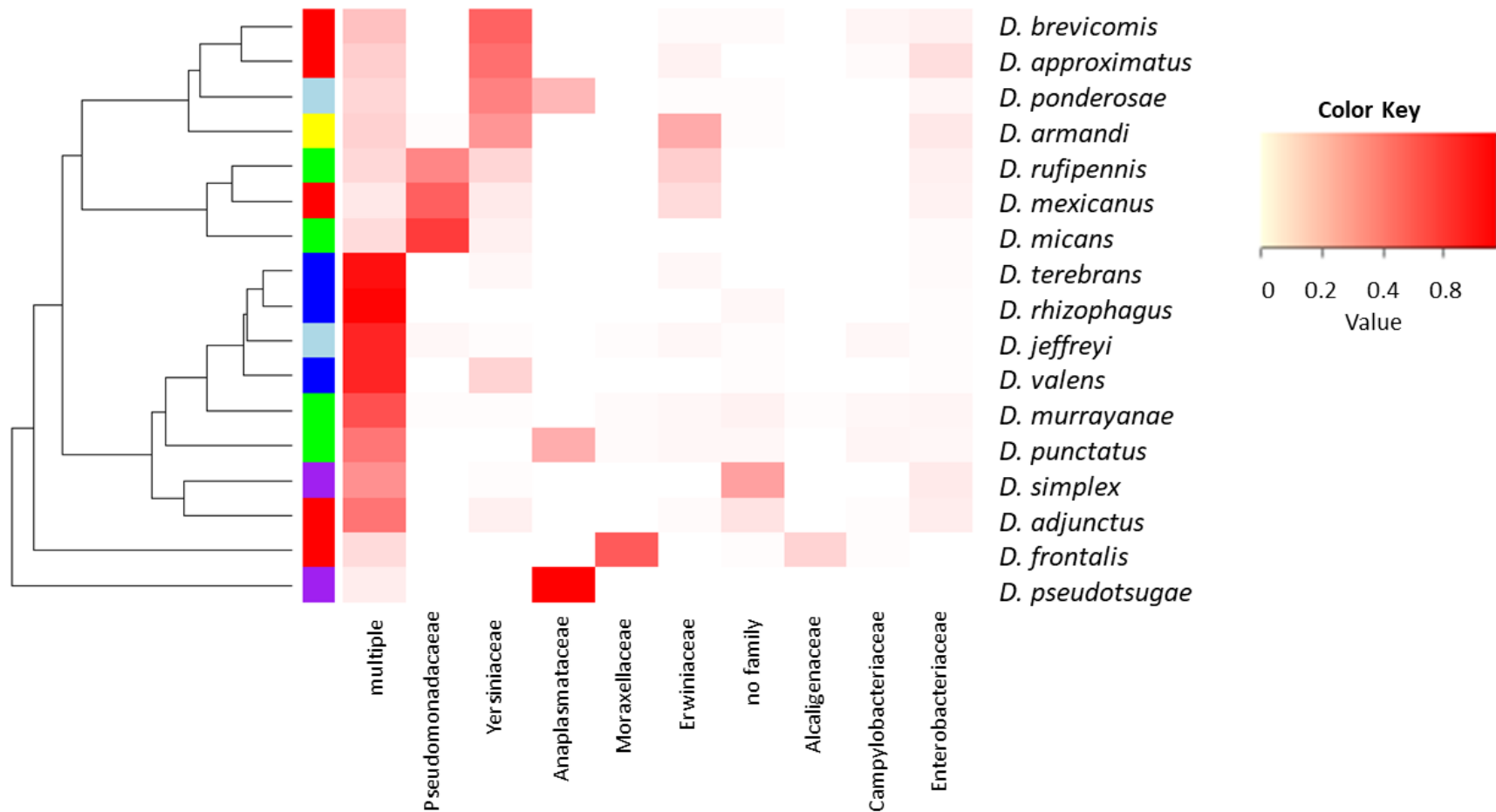
En algunos casos como *D. pseudotsugae* y *D. murrayana* sólo se reportó la presencia de un loci de todos los taxa.

Entre los 5,543 loci de Proteobacteria recuperados por *Denovo\_97*, se encontraron 10 categorías de familia con una abundancia relativa de más del 1% en al menos una de las especies consideradas (Fig. 17). La familia más numerosa fue Pseudomonadaceae con un total de 5,104 secuencias homólogas distribuidas entre las diferentes muestras, seguido de Yerseniaceae con 2,590, Anaplasmataceae con 2,244 y Moraxellaceae con 1,121.

Cabe señalar que la categoría “multiple” superó al resto de las familias reales con un total de 4,570 secuencias homólogas distribuidas entre todas las especies de escarabajo. Además, el valor máximo de esta categoría fue de 1,140 secuencias homólogas encontrados en las muestras de *D. valens* mientras que el mínimo fue de 49 secuencias homólogas en *D. simplex*. Otra categoría que resaltó en este análisis fue “no family”, en la cual se encontraron 245 secuencias homólogas distribuidas también entre todas las especies del género. Estas cifras son de gran importancia para este estudio dada la gran cantidad de loci bacterianos reportadas exitosamente hasta nivel de phyla, pero difícilmente clasificados en niveles taxonómicos más precisos.

Al analizar el dendograma construido a partir de las abundancias relativas calculadas para cada especie, se encontraron dos pares de especies en las que se conservan de manera incompleta los grupos filogenéticos considerados en la Fig. 3. Los nodos formados por *D. brevicomis* y *D. approximatus* (grupo III) y por *D. rhizophagus* y *D. terebrans* (grupo I) destacan entre el resto del árbol, además de que estas últimas especies del grupo I se encuentra muy próximas a su especie hermana *D. valens*, separadas únicamente por la inclusión de *D. jeffreyi* (grupo III) que propone la topología.

### Distribución de secuencias homólogas de las principales familias de Proteobacteria de *Denovo\_97*



**Figura 17.** Mapa de calor de las principales familias de Proteobacteria recuperadas en cada especie de *Dendroctonus* a partir del ensamble *Denovo\_97*. El gradiente de color de cada phyla relacionado con una especie de escarabajo corresponde a su abundancia relativa. Se descartaron a aquellos phyla que no presentaban una abundancia relativa mayor al 1% en al menos una especie hospedera. Se construyó un dendrograma basándose en los patrones de asociación de los phyla

informativos entre las especies de *Dendroctonus* mediante el método de agrupación jerárquica. Estas especies se etiquetaron con los colores asociados a los grupos filogenéticos de la Fig. 6: I Amarillo, II Morado, III Verde, IV Azul claro, V Rojo y VI Azul.

Se mapeó la distribución de 2,161 loci clasificables hasta nivel de género recuperados por *Denovo\_97* pertenecientes a múltiples phyla informativos: Nematoda, Proteobacteria, Streptophyta, Ascomycota y Microsporidia (Fig 19). Dentro de Proteobacteria, el phylum más abundante, los géneros con la mayor cantidad de loci fueron: *Pseudomonas*, *Wolbachia* y *Rahnella* con 608, 449 y 182 respectivamente.

Al analizar a profundidad la determinación taxonómica realizada por este método (Anexo 8), se encontró que en el género *Pseudomonas* se identificaron 41 especies bacterianas diferentes con más de 97% de similitud, de las cuales la mayor parte de los loci se distribuyeron entre las muestras de *D. ruffipennis* y *D. micans* (550 y 552 loci), seguido de *D. mexicanus* (126 loci). Dentro del género *Wolbachia* se identificaron múltiples representantes de los subclados A y B, responsables de la infección de invertebrados y de patologías como la incompatibilidad citogenética o la inducción de la partenogénesis (Werren, 1997). La gran mayoría de las secuencias homólogas de este género de importancia biológica (1,734) se concentraron en *D. pseudotsugae*, con sólo menos del 15% (250 y 40) asociados a *D. ponderosae* y *D. punctatus*.

Algunos otros géneros de interés recuperados con más de 97% de similitud incluyen 124 loci de *Pinus* repartidos entre *D. terebrans* (52), *D. murrayanae* (28), *D. micans* (20), *D. valens* (12) y *D. approximatus* (4). También se encontraron 45 loci del género de pináceas *Picea* distribuido entre *D. micans* (13), *D. murrayanae* (8), *D. terebrans* (10), *D. punctatus* (4), *D. ruffipennis* (4), *D. valens* (3) y *D. approximatus* (3).

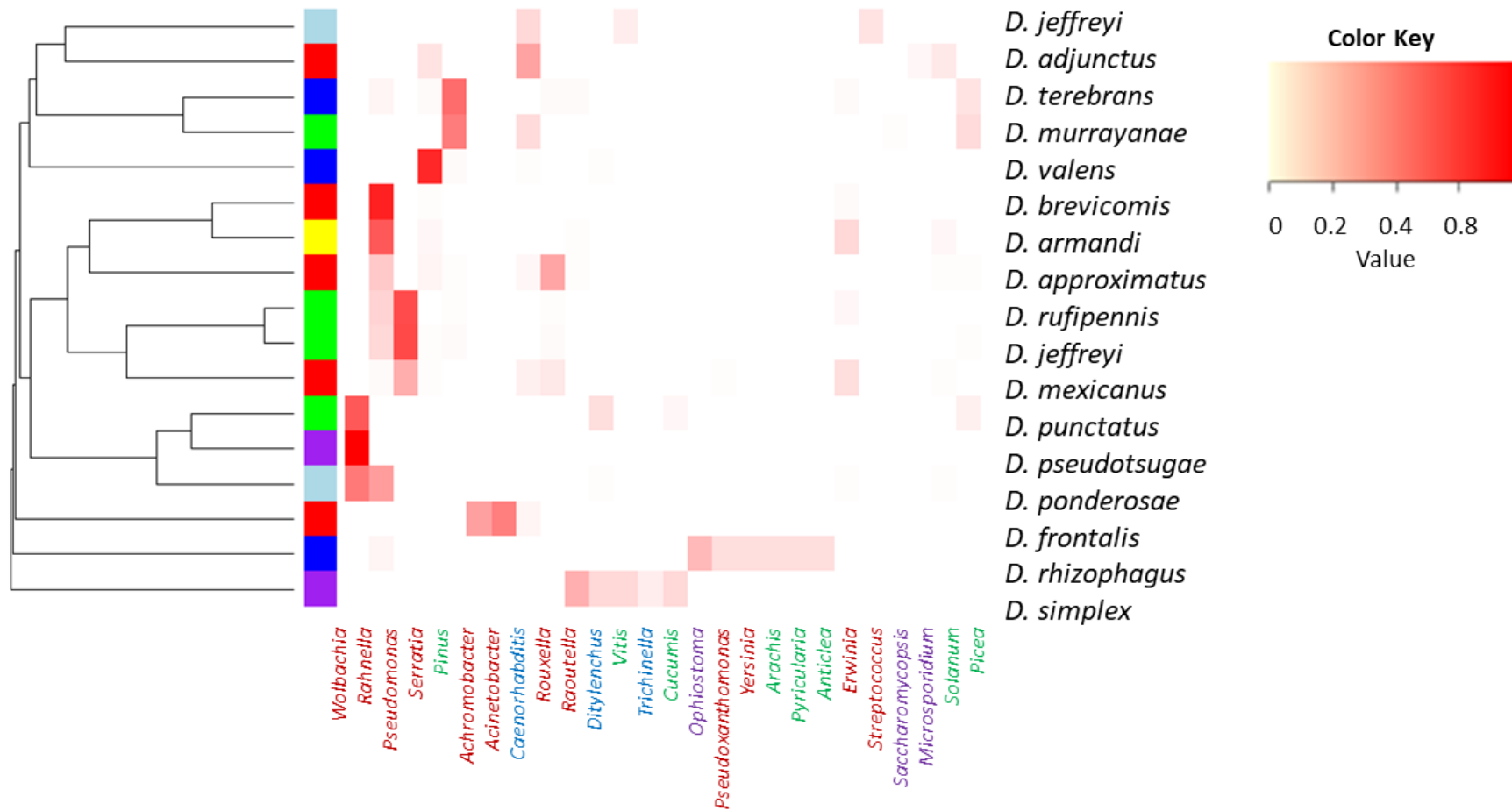
Además, se encontró un loci único formada por tres secuencias homólogas del género de levadura *Saccharomycopsis* (Saccharomycetaceae), distribuidos entre las especies *D. micans*, *D. marroyanae* y *D. ruffipennis*, así como cuatro loci del género *Microsporidium* reportados únicamente en *D. adjunctus*.



Entre los 41 loci con más de 97% de similitud pertenecientes a 5 géneros de Nemátodos, el más abundante fue *Caenorhabditis* (26). A partir de los 26 loci se encontraron 162 secuencias homólogas distribuidas en *D. frontalis*, *D. adjunctus*, *D. mexicanus*, *D. approximatus*, *D. murrayanae*, *D. valens*, *D. ponderosae*, *D. pseudotsuga* y *D. jeffreyi*.

Las únicas especies que presentan un patrón de asociación en el que se conservan las relaciones filogenéticas fueron *D. micans* y *D. ruffipenis* (Gpo. III). Este grupo parece haberse construido debido a las semejanzas entre las abundancias relativas que se asumen de géneros como *Pseudomonas*, *Serratia* y *Picea* (Fig. 18).

### Distribución de secuencias homólogas de los géneros recuperados por *Denovo\_97*



**Figura 18.** Mapa de calor de los géneros recuperados mediante el ensamble *Denovo\_97*, con más de 97% de similitud, pertenecientes a phyla informativos con una abundancia relativa mayor al 1% en al menos una muestra y su distribución en cada especie de *Dendroctonus*. Se construyó un dendrograma basándose en los

patrones de asociación de los phyla informativos entre las especies de *Dendroctonus* mediante el método de agrupación jerárquica. Estas especies se etiquetaron con los colores asociados a los grupos filogenéticos de la Fig. 6: I Amarillo, II Morado, III Verde, IV Azul claro, V Rojo y VI Azul. Los géneros encontrados se colorearon de acuerdo al reino al que pertenecen. Rojo: Proteobacteria, Verde: Plantae, Azul: Animalia, Morado: Fungi

## DISCUSIÓN

Se eligió al género *Dendroctonus* como modelo para evaluar el alcance de la secuenciación de tipo RAD debido al papel que han tenido estos organismos en el estudio de la simbiosis de insectos con otros grupos como hongos, bacterias, nemátodos, ácaros y plantas. Además, la relevancia de sustentar este trabajo con secuencias provenientes de este género de escarabajos radica en el impacto socioambiental que tienen a lo largo del continente, en particular algunas especies menos estudiadas. Si bien, los brotes de especies como *D. ponderosae* se han considerado como una amenaza ambiental prioritaria para millones de hectáreas de pinares de Canadá y Estados Unidos (Vore *et al.*, 2020), las especies con un mayor impacto en México y América Central no sólo afectan de manera directa a la salud de las comunidades forestales, sino que la gestión ineficiente de los brotes ha aumentado la frecuencia y la intensidad de los incendios forestales. El empleo de la quema como estrategia de manejo ambiental ha impedido la regeneración natural en zonas donde estas tierras a menudo se aprovechan para usos agrícolas o silvícolas (Billings *et al.*, 2004).

La capacidad de realizar sondeos de la microbiota de hospederos poco estudiados utilizando datos de secuenciación de RAD, puede tener un papel clave para abrir el camino a líneas de investigación enfocadas en algunos grupos de simbiosis o en la distribución de estos a lo largo de la filogenia del escarabajo. Sin embargo, es necesario reconocer los límites de este método y de la interpretación de sus resultados. En la discusión de este trabajo se abordan algunas consideraciones oportunas para el desarrollo de esta técnica y para relacionar este análisis con el contexto biológico de los escarabajos descortezadores del género *Dendroctonus*.

### Consideraciones metodológicas

#### Modelo biológico y diseño experimental

Para abordar correctamente los resultados obtenidos, es necesario profundizar en las condiciones bajo las cuales se generó la librería de RADSeq para el género *Dendroctonus* empleada en este proyecto. Al haberse generado originalmente para un estudio filogenético sobre este género, se ocuparon individuos provenientes de múltiples localidades de Norteamérica y Eurasia, de los cuales se colectaron al menos tres individuos de cada especie, por lo que no existió un control estricto de variables biológicas y ambientales como podría desearse para este tipo de sondeos metagenómicos (Godefroid *et al.*, 2019; Alberdi *et al.*, 2021).

Tras indagar en el origen de los datos empleados para este estudio (Godefroid *et al.*, 2019) la primera variable que surge a discusión para la optimización de un sondeo metagenómico es la inclusión de diecisiete especies de *Dendroctonus* y la población de la que estas provienen. Se ha discutido que la divergencia genética entre especies puede tener un impacto en las interacciones entre el genoma de las especies hospederas con su microbiota (Alberdi *et al.*, 2020). Esto puede parecer alarmante debido a la variación de la fórmula cariológica dentro de este género (Zúñiga *et al.*, 2009), sin embargo, este impacto puede variar de acuerdo a la resolución del sondeo (Alberdi *et al.*, 2020). Mientras se definan los objetivos del proyecto de manera concisa es posible estudiar a las comunidades compartidas entre diferentes especies, conocida como microbioma central, el cuál como ya se ha mencionado ha sido previamente estudiado en *Dendroctonus* (Hernández-García *et al.*, 2017).

Al considerar la población de origen, se destaca el impacto que pueden tener múltiples variables ecológicas asociadas a la misma. En el caso de los datos empleados para este trabajo, los ejemplares de especies como *D. jeffreyi*, *D. armandii* y *D. rhizophagus* fueron muestreados de una misma localidad por cada especie, sin embargo la gran mayoría de las especies incluyeron individuos provenientes de diferentes poblaciones, sin un control aparente de la época del año o de la etapa pre o post-emergente de su pino-hábitat (Anexo 9). En principio esta decisión del muestreo resulta lógica debido a los objetivos originales del estudio, pero esta resulta importante tomar en cuenta el efecto que tiene el estadio del ciclo de vida en la diversidad de la microbiota intestinal encontrada en especies como *D. rhizophagus* y *D. armandi* (Briones-Roblero *et al.*, 2017; Hu *et al.*, 2012), o la época del año y la temperatura en la estructura de la microbiota intestinal en *D. valens* y *D. ponderosae* (Hou *et al.*, 2021; Fraser *et al.*, 2011).

De igual forma se ha demostrado que tanto la composición de la comunidad microbiana como su perfil ecológico-funcional no es una entidad estática a lo largo de la vida de los individuos. La microbiota se encuentra relacionada con factores como el fenotipo y la dieta del hospedero, o bien, con las condiciones ambientales (Alberdi, *et al.*, 2021). Al emplear una librería de RADSeq construida a partir de muestras de diferentes poblaciones, es posible que las condiciones ambientales a las que estaban sujetas tengan una influencia importante en la resolución de la determinación taxonómica, es decir, en la capacidad de identificar a las secuencias ensambladas hasta niveles taxonómicos específicos. A partir de esto se puede inferir que debido al método y al modelo biológico empleado, muy posiblemente no se cuenta con el alcance para denotar las diferencias entre la microbiota de las poblaciones, pero sí puede ser utilizado para recuperar información del microbioma central compartido entre las diecisiete especies del género.

## Ensamble y recuperación de loci

Una de las principales razones por la cual el método que propone este trabajo se diferencia de las opciones actualmente disponibles para el estudio de la microbiota, radica en la naturaleza estocástica de la digestión enzimática que caracteriza a los datos de RADSeq (Holmes y Rabosky, 2018). Esta misma propiedad, le permite recuperar lecturas asociados a taxa procariotas y eucariotas (unicelulares y multicelulares), sin depender de la selección de amplicones para grupos particulares, sin embargo la presencia de microorganismos de diferente phyla, representa un reto para su asignación taxonómica.

El flujo de trabajo adoptado para el presente trabajo nos permitió comparar e identificar las ventajas y desventajas de los métodos de ensamble mediante el pipeline de ipyrad. Es posible observar esto en el impacto que tiene el valor del parámetro CT en las diferentes aproximaciones evaluadas. Los ensambles + *referencia* con un valor de *mindepth\_majrule* de 2, recuperan una mayor cantidad de loci al disminuir el valor de CT a 0.90 y 0.85, debido a que al disminuir este valor se permite que los loci con menor similitud con respecto a la secuencia de referencia sean recuperados por el ensamble. En el caso de *GRef\_85\_m2* es posible explicar los resultados positivos de la disminución de este parámetro, al tomar en cuenta que en este estudio se realiza el sondeo a partir de secuencias de especies diferentes y se emplea una secuencia de referencia para la búsqueda, por lo que mientras más laxo sea el parámetro de similitud empleado, más probable es que se recuperen loci que presentan un mayor número de pares de bases diferentes a la secuencia de referencia.

En contraste, en los ensambles *de novo* y *de novo - referencia* se observa una tendencia opuesta, en la que al emplear valores de CT mayores se recupera una mayor cantidad de loci. A pesar de que mediante ensayos empíricos se ha observado que un valor de CT alto suele relacionarse con la pérdida de loci (McCartney-Melstad *et al.*, 2019), el aumento en el número de loci recuperados con el valor más alto de CT puede estar relacionado con la separación de las secuencias que presentan suficiente divergencia para ser consideradas loci independientes, en lugar de alelos del mismo loci. El resultado de este fenómeno es un mayor número de loci, los cuales no son descartados debido a la modificación de otros parámetros como el *min\_sample\_per\_locus* con valor de 1.

Otro de los parámetros que cobraron importancia en este ensayo fue *mindepth\_majrule*, cuya disminución se derivó de la comparación de los estadísticos de resumen de los ensambles + *referencia* (desglosados en el archivo *.stats* que ipyrad genera como output), en el que se identificó una cantidad anómala de loci excluidos por los filtros del programa. Posteriormente se estudió el mecanismo de filtrado que se realiza durante los diferentes pasos del programa, destacando el paso 5, durante el cual se utiliza un criterio de recuperación de cada base que depende de un valor de profundidad (*base*

*calling*). Se sugirió que esta profundidad límite, determinada por los valores predeterminados de los parámetros, podría no ser superada por los loci de las secuencias analizadas debido a la subrepresentación del DNA microbiano.

Al comparar los resultados de los ensamblajes *GRef* con el parámetro *mindepth\_majrule* modificado, el impacto positivo de la disminución del valor a 2 parece demostrar la subrepresentación que tienen los loci microbianos en el set de datos. Este fenómeno puede explicarse como una consecuencia de la técnica empleada, ya que al preparar las librerías genómicas de RADSeq de manera estándar, se parte de una mayor disponibilidad de DNA de los hospederos con respecto a la de su microbiota, por lo que las lecturas microbianas tienden a tener una profundidad mucho menor (Pereira-Marques *et al.*, 2019). Además, esto cobra importancia en el área, debido al impacto de la profundidad de secuenciación en la cobertura del metagenoma, entendida como el porcentaje de información metagenómica recuperada de la comunidad microbiana, por lo que debe ser considerada con el fin de generar una descripción más acorde a las comunidades reales (Sims *et al.*, 2014; Pearman *et al.*, 2020).

El ensamblaje *Denovo\_97* se estableció como el modelo para la exploración de algunos grupos de interés biológicos debido al éxito reflejado en la cantidad de loci y de taxa diferentes que recuperó. Por su parte los ensamblajes denotados como *DRef* generaron resultados similares, sin alcanzar la misma cantidad de loci microbianos o de interés biológico que el método de *Denovo*. Si bien el método *de novo - referencia* presenta la ventaja de filtrar a aquellos loci similares a la referencia, es decir, a un gran porcentaje de los loci de Arthropoda que con otros métodos de ensamblaje se descartan hasta etapas posteriores del análisis (Fig 6), es posible que un porcentaje de los loci de interés biológico se pierda desde el ensamblaje de loci en ipyrad, por lo que se realizó un análisis de BLAST de la secuencia del genoma completo de *D. ponderosa* que se empleó como referencia.

Del análisis de la secuencia del genoma *draft* de *D. ponderosa* usada como referencia (Fig. 7), se reporta la posible influencia de secuencias asociadas a simbiosis o posibles contaminaciones. Aunque las secuencias fúngicas encontradas en la secuencia de referencia pueden relacionarse con la incompletitud del genoma publicado, esto debe tomarse en cuenta para prevenir que se filtre a una parte de la comunidad microbiana de interés durante el proceso de ensamblaje. Cabe destacar que entre los casi 300 loci fúngicos encontrados con más de 90% de similitud, únicamente uno de estos registró un nivel mayor al 97% de similitud (Anexo 7), el cual corresponde al género *Fusarium*, que se caracteriza por su abundancia en ecosistemas forestales, por su hábito saprófito entre la microbiota del suelo y por incluir especies patógenas de plantas y animales (Ma *et al.*, 2013). Dado este resultado, se confirma que la naturaleza de la secuencia de referencia utilizada en este tipo de ensamblajes puede

tener una influencia contraproducente para el sondeo de la microbiota si esta no mantiene una integridad taxonómica certera, por lo que es importante considerar las condiciones en las que se generan las secuencias genómicas. Esta es un área de oportunidad para realizar estudios enfocados hacia ciertos simbioses u hospederos que se desarrollará conforme se publiquen una mayor cantidad de secuencias bien trabajadas. Incluso, de continuar desarrollándose este método y publicándose más secuencias genómicas de diversos grupos, podrían realizarse ensambles + *referencia* de manera sucesiva a partir de los taxa encontrados en cada uno de estos, para dirigir aún más cada búsqueda y diseñar ensayos posteriores que confirmen la presencia de los loci de interés.

En general, los ensambles *Denovo - Referencia* recuperan menos loci que aquellos realizados mediante el método *Denovo* (Fig 6) debido a la naturaleza de este ensamble, sin embargo *DRef\_95* recupera una mayor cantidad y diversidad de loci informativos y no informativos que *Denovo\_95*. Este fenómeno no se observa en los ensambles con un CT de 0.97 (*Denovo* y *DRef*), por lo que esto puede indicar que 0.95 es un valor de CT ideal cuando se emplea la referencia del hospedero para filtrar. Por otra parte el valor de CT de 0.97 resultó crítico en los ensambles *Denovo*, presentando una diferencia de un orden de magnitud entre la cantidad de loci recuperados, este impacto demuestra la sensibilidad que tienen algunos set de datos a los parámetros de ipyrad y la utilidad de un método empírico adaptable a las condiciones de cada set de datos de RADSeq, del modelo y de los objetivos de cada sondeo (McCartney-Melstad *et al.*, 2019).

Por su parte los ensambles *DPond*, realizados con el método *denovo - referencia* y las muestras de la especie de la secuencia genómica empleada como referencia (*D. ponderosae*), demuestran las limitaciones de restringir un estudio a una menor cantidad de muestras de una misma especie en contraste con múltiples muestras de diferentes especies. Las ventajas de priorizar el número de muestras sobre la identidad de estas, han sido previamente reportadas en el ensamble de lecturas metagenómicas para la identificación de bacterias (Jovel *et al.*, 2016), lo cual sugiere que este efecto se presenta también con los datos de RADSeq.

## **Determinación taxonómica**

Uno de los pasos clave de este método consistió en seleccionar aquellas secuencias con más de 90% de similitud con respecto a una base de datos, para su análisis posterior. Partiendo de la cantidad de loci de phyla no informativos encontrados por el ensamble *Denovo\_97*, es lógico cuestionarse cómo explicar el porcentaje de coincidencias con phyla que no presentan una relación con el contexto biológico de los hospederos. Aunque es posible relacionar la presencia de loci de Chordata con el



sesgo de la bases de datos nucleotídica empleada o incluso con posibles contaminaciones antrópicas de las muestras, la presencia de otros phyla descartados como Cnidaria, Echinodermata, Mollusca o Porifera nos lleva a preguntarnos cómo influye el tipo de datos y las herramientas que propone este método. Partiendo de los criterios más laxos empleados en los estudios de la microfauna y la mesofauna asociada al suelo, es posible descartar a los phyla que se caracterizan por presentar un tamaño de más de 2 mm por su poca probabilidad de encontrarse asociados a las especies de *Dendroctonus* de menor tamaño, o bien por no habitar en ambientes terrestres como *Dendroctonus* (Kampicher *et al.*, 2009; Armendáriz-Toledano *et al.*, 2018).

La dificultad asociada a la asignación taxonómica ha sido atribuida al sesgo de las bases de datos empleadas, cuyo crecimiento depende de las tendencias de estudio en la ecología microbiana. A menudo los análisis metagenómicos, los cuales no dependen de la selección de amplicones, se limitan únicamente a la diversidad microbiana asociada a una muestra o incluso a la diversidad de organismos procariontes. Esto puede relacionarse con el mayor número de diversidad calculada para estos grupos, la cantidad considerable de taxa microbianos desconocidas por su incapacidad de ser cultivadas, la capacidad de cuantificar gran parte de la diversidad eucariota mediante métodos tradicionales, o la escasez de información genómica eucarionte disponible para este tipo de estudios (Escobar-Zepeda *et al.*, 2015).

Además, se ha demostrado que el tamaño de las lecturas metagenómicas empleadas para la asignación taxonómica puede tener una influencia significativa dependiendo del grupo que se busque estudiar (Pearman *et al.*, 2020). Esto representa una desventaja importante para el aprovechamiento de la información de RADSeq debido al tamaño de los loci generados (~100 bp), ya que a pesar de que es posible seleccionar un tamaño mínimo de las lecturas ensambladas entre los parámetros de ipyrad, se prevé que esto puede reducir la cantidad de loci recuperados.

A medida que se ha desarrollado la metagenómica como una rama científica y una herramienta fundamental en la ecología microbiana, han surgido también diferentes plataformas de secuenciación que han permitido estudiar la influencia de propiedades técnicas como la longitud de lecturas generadas. Los reads de tamaño más cortos, regularmente asociados a las plataformas de secuenciación de segunda generación (~150 bp en pirosecuenciación), no son recomendados para la caracterización de comunidades microbianas particularmente por la variabilidad de ciertos taxa en las bases de datos (Wommack *et al.*, 2008). Mientras que aquellos de tamaño mayor (125 - 300 bp en plataformas Illumina) han demostrado tener un mejor desempeño para la clasificación de bacterias que para los de taxa eucariotas. En contraste con esto, las lecturas de gran tamaño asociados a plataformas de secuenciación de última generación (4 Kbp en Nanopore), a pesar de estar más sujetos

a errores, son más útiles para lograr una clasificación exitosa de comunidades eucariotas e incluso han sido señalados por su potencial utilidad para realizar clasificaciones a nivel de cepa o de linajes subespecíficos en comunidades bacterianas (Pearman *et al.*, 2020; Maguire *et al.*, 2021).

La influencia que puede tener el tamaño de las lecturas dependiendo del grupo de estudio puede ser explicado por múltiples factores, como puede ser el hecho de que los taxa bacterianos presentan un menor grado de divergencia genética (aprox. 6% de identidad nucleotídica promedio, ANI por sus siglas en inglés, a nivel genómico), mientras que en los grupos eucariotas no se ha logrado definir más que en algunos loci de bajo nivel de divergencia, como es el caso del gen COI mitocondrial (Pearman *et al.*, 2020). Otro factor de gran importancia se refiere a la cantidad de elementos repetitivos que es posible encontrar en plantas y animales a comparación de la relativa simplicidad de los genomas fúngicos y procariotas, estos elementos repetitivos se encuentran distribuidos entre diferentes taxa de estos dos reinos de manera muy conservada, por lo que suelen ser “ignorados” por los algoritmos de los clasificadores, sin embargo, al emplear reads de tamaño corto y de origen aleatorio como los de RADSeq, estos algoritmos son más propensos a fallar (Pearman *et al.*, 2020).

Al analizar los tamaños de las secuencias con las que se realizó la determinación taxonómica de los dos principales ensamblajes de este ensayo, *Denovo\_97* y *GRef\_85\_m2*, se encontró una tendencia hacia el valor de 97 pb con poca diferencia entre la desviación estándar de los diferentes taxa en ambos grupos. Este resultado se relaciona con el método bajo el cual se construyó la librería de RADSeq originalmente, el cual empleó un paso de sonicación mecánica y selección en gel para posteriormente realizar la secuenciación pareada de ~125 pb. Si se toman en cuenta las bases secuenciadas que corresponden a los adaptadores y los identificadores de las muestras, es posible explicar las tendencias observadas en el tamaño de los loci de los diferentes taxa. Los únicos loci con tamaño promedio anómalo que se detectaron pertenecen a Apicomplexa y Chlorophyta, recuperados por el ensamblaje *GRef\_85\_m2*, pero dada su abundancia reportada de 2 o 1, pueden ser estadísticamente descartables de este y futuros ensayos. A partir de esto se propone establecer un criterio de discriminación a partir de la abundancia o tamaño de loci para futuros sondeos metagenómicos como este.

Cabe señalar que el trabajo de Pearman y colaboradores de 2020, profundiza en el impacto que puede tener el tamaño de las lecturas para la asignación taxonómica comparando las tecnologías de secuenciación de segunda y tercera generación disponibles en la actualidad, con diferentes softwares dedicados a esta tarea. A partir de estos análisis se concluye que el tamaño de las lecturas tiene una influencia variable dependiendo del nivel taxonómico y del grupo que se busque identificar,

particularmente la diversidad eucariota de plantas y animales presenta mayores problemas para ser identificada a nivel de género a partir de lecturas cortas. Para bacterias y hongos, la recuperación aumentó casi 80% cuando se usan lecturas de 1500 pb sobre las de 100 pb, mientras que para animales y plantas se observaron tendencias similares con tamaños de lectura un poco más grandes, de 3000 pb para animales y 2500 pb para plantas, sin importar del software clasificador empleado.

Además, se encontró que en el caso de los reads de tamaño corto (~100 pb) de animales y plantas, el software de asignación taxonómica Kraken 2 tiene un mejor desempeño que BLAST, posiblemente debido a los parámetros sub-óptimos que se ha reportado que puede presentar este programa con los fragmentos de tamaño pequeño (Pearman *et al.*, 2020).

### **Taxa raros**

Al observar la diversidad taxonómica obtenida por medio del ensamble más exitoso, *Denovo\_97* (figura 8), se encuentran 67 categorías de phyla diferentes. Entre los phyla anómalos que se descartaron por su poca relación con el contexto biológico de las especies de *Dendroctonus* destacan aquellos mayoritariamente asociados a ambientes marinos, sin embargo, incluso entre los phyla informativos es difícil establecer una relación directa entre los grupos encontrados y la mayor parte de los loci asignados a familias u otros taxa.

En el caso de los loci asignados como plantas, bacterias y hongos se observa un fenómeno análogo: la presencia de decenas de familias *raras* o bien, subrepresentadas entre los loci recuperados por los diferentes ensambles. Algunos de los factores que se considera que tiene una influencia importante en este resultado es el tamaño de los fragmentos de RADSeq y la composición de las bases de datos empleadas en este trabajo (Pearman *et al.*, 2020).

Otras técnicas de sondeo metagenómico como el *metabarcoding* han superado algunos retos similares para caracterizar correctamente a los *taxa raros* mediante métodos de normalización o enriquecimiento al construir las librerías. El estudio de estos taxa representa un reto para comprender la dinámica de determinadas comunidades microbianas e identificar su posible función en estas mismas (Hildebrand, 2021), sin embargo, es preciso señalar que los taxa mencionados como “raros” encontrados en las secuencias de RADSeq de *Dendroctonus*, debido a la naturaleza heterogénea de muestreo, no pueden ser explicados ni tratados con el mismo enfoque. Es posible que en el trabajo de Holmes y Rabosky, enfocado únicamente en secuencias obtenidas de tejidos del hospedero, no se haya enfrentado a una diversidad tan grande de taxa al tener un mayor control en el muestreo.

Aunque no se hayan encontrado diferencias entre el tamaño de las secuencias clasificadas como “taxa raros” y aquellas que mantienen una relación con el contexto biológico del género *Dendroctonus*, o bien, pertenecen al propio escarabajo hospedero, es posible atribuir muchos de los resultados inconclusos de la asignación taxonómica al tamaño de los loci y a la naturaleza de la base de datos empleada. Tomando en cuenta la presencia de elementos repetitivos compartidos entre múltiples genomas eucariotas, particularmente de Plantas y Animales, y los tamaños de menos de 100 pb de las secuencias empleadas para la asignación taxonómica (Tabla 4), es posible sugerir que los taxa de estos dos reinos tienden a ser menos confiables, en especial cuando se busca llegar a niveles taxonómicos de familia, género o especie (Pearman *et al.*, 2020).

## **Comparaciones de diversidad**

A lo largo de este trabajo se ha resaltado el potencial que esta metodología puede tener para realizar sondeos de la diversidad microbiana de manera más accesible o incluso inmediata, a partir de datos previamente publicados, no obstante, es importante ahondar en los supuestos asumidos al cuantificar e interpretar los resultados obtenidos. La secuenciación de RADSeq posee inherentemente un componente estocástico porque depende de digestiones enzimáticas cuya secuencia de corte suele ser pequeña y común, para así obtener la mayor cantidad de loci posibles (Andrews, 2016, Schweyen *et al.*, 2014). El presente método se aprovecha de esta condición para identificar coincidencias con simbiontes microbianos o con integrantes de la comunidad del hospedero (Holmes y Rabosky, 2018), por lo que al comparar las abundancias de ciertos loci en las diferentes especies de escarabajo y al proponer relaciones ecológicas con los taxa encontrados, se conserva el valor exploratorio de estos resultados.

De igual manera, estas circunstancias también fueron determinantes para reconocer los alcances del análisis. Las implicaciones de intentar describir a las comunidades ecológicas o de proponer posibles eventos de coevolución a partir de la cantidad de loci reportado en cada muestra de escarabajo son profundas. El control de las condiciones asociadas con las técnicas de biología molecular y los análisis bioinformáticos puede parecer sencillo, pero se sugiere que este depende profundamente del desarrollo de este método en sus diferentes etapas.

## **Consideraciones ecológicas**

Entre los múltiples phyla encontrados destacan algunos por su cantidad y por su relevancia en el acervo biológico reportado de las diferentes especies de hospederos. Al concentrarnos en los

ensambles más exitosos, observamos diferencias en la cantidad de loci asignados a determinados grupos seleccionados por la relación biológica que mantienen con *Dendroctonus*.

## Fungi

En cuanto a la relación simbiótica entre hongos y escarabajos, se destacaron dos métodos de ensamble: *Denovo* por haber reportado una mayor diversidad de taxa (todos los phyla que integran a el reino Fungi a excepción de Glomeromycota) y a *GRef* con el parámetro *mindepth\_majrule* disminuido a 2, por la gran cantidad de loci fúngicos que recuperó. Además, la mayor parte de los loci recuperados por este último método coincidieron con la especie de la secuencia de referencia empleada: *Grosmannia clavigera*. Este resultado concuerda con las condiciones del método, el cual limita la búsqueda a aquellos loci con mayor similitud a la secuencia de referencia, además de indicar que los loci fúngicos están presentes a un nivel bajo de profundidad de secuenciación. Como se discutió anteriormente, es posible que estos loci provengan de otros simbiontes micangiales relacionados filogenéticamente con *G. clavigera*, sin embargo, al ser esta la única especie de la familia Ophiostomataceae cuyo genoma ha sido publicado en la base de datos del NCBI, la mayor parte de las coincidencias son representadas por ella.

Al analizar la distribución de loci encontrados por *GRef\_85\_m2*, se encontró que la mayor parte de los loci fúngicos, pertenecientes a Ophiostomataceae, se distribuían entre *D. jeffreyi*, *D. mexicanus* y *D. brevicomis* (Fig. 16). Estas últimas dos especies presentan micangio protorácico, mientras que *D. jeffreyi* y su grupo filogenético, caracterizados por presentar un micangio maxilar, se caracteriza por albergar específicamente a *G. clavigera* (DiGustini, 2019).

Las otras nueve familias fúngicas encontradas por *GRef\_85\_m2* pueden parecer descartables debido al bajo número de loci recuperados y la cantidad de muestras de escarabajo en donde se encontraron, sin embargo la posibilidad de encontrar coincidencias con especies simbiontes cuya secuencia genómica se encuentre publicada, puede abrir la puerta a realizar más ensambles con el método + *referencia*. Estos ensambles posteriores pueden ser de utilidad para localizar nuevos loci asociados a esta especie de referencia en otros hospederos o bien, para expandir el estudio y la diversidad de taxa sondeados. Dicho proceso de ensambles “en cadena” podría ser viable en un futuro, tanto por el bajo costo computacional, como por el crecimiento acelerado de las bases de datos de referencia.

Del total de secuencias agrupadas por *Denovo\_97* sólo el 0.11% (216 loci) se identificaron como loci fúngicos. Al igual que en los resultados anteriores, nuevamente se identificó a Ascomycota como el phylum más numeroso de este reino, seguido por Basidiomycota y Mucoromycota. La familia de ascomicetos simbioses obligados reportada previamente en casi la mitad de las especies de *Dendroctonus* estudiadas, Ophiostomataceae, se encuentra subrepresentada ante otras familias como Saccharomycetaceae. Mientras que de Ophiostomataceae se encontraron sólo 2 loci, de Saccharomycetaceae se encontraron 31, esto puede estar relacionado una vez más con el aprovechamiento que tienen algunas especies de Saccharomycetaceae en la industria biotecnológica y la influencia que puede tener esto en su estudio y en la publicación de sus secuencias en la base de datos del NCBI.

Entre los phyla fúngicos encontrados por *Denovo\_97* se reportó a Ascomycota como el más abundante de estos, cuya distribución se concentró principalmente en *D. frontalis* con 67 secuencias homólogas, *D. micans* con 60 y *D. mexicanus* con 59. A pesar de que dos de estas especies presentan micangio protorácico (*D. frontalis* y *D. mexicanus*), al explorar los géneros con más de 97% de similitud encontrados por *Denovo\_97*, sólo se encontraron 12 secuencias homólogas correspondientes al género Ophiostoma (Fig. 17), el cual incluye a algunas especies de simbioses micangiales, distribuidas en su totalidad en este ensayo en *D. ruffipenis*.

El único representante de Saccharomycetaceae con un porcentaje de similitud de más de 97% encontrado fue *Saccharomycopsis* (Fig. 17), representado por 3 secuencias homólogas en *D. ponderosae*, *D. murrayanae* y *D. ruffipenis*. Este género ha sido previamente reportado en *D. ponderosae*, *D. valens* y *D. simplex* (Lou *et al.*, 2014; PO *et al.*, 2006; Durand, 2018) y se cree que puede mantener interacciones funcionales con su hospedero tales como la degradación de compuestos vegetales complejos o el metabolismo de hormonas relacionadas con la atracción de escarabajos de la misma especie o la repulsión de posibles competidores (Briones Roblero *et al.*, 2016).

También se reportó la presencia de 4 secuencias homólogas en el género *Microsporidium* en *D. adjunctus*, el cual si bien no ha sido reportado previamente en este género, sí se han encontrado en *D. micans*, *D. pseudotsugae* y *D. frontalis* como patógenos intracelulares de estas especies cuyo impacto puede llegar a variar mucho dependiendo de la población y la especie de escarabajo infectada (Yaman *et al.*, 2008; Weiser, 1970, Knell y Allen, 1978). No se encontraron géneros de Basidiomycota con al menos 97% de similitud.

## Plantae

Al explorar el phylum Streptophyta que incluye a plantas vasculares y no vasculares, con el fin de evaluar la presencia de DNA asociado a la ingestión de tejido vegetal de pináceas por la cual es conocido este género, se encontraron resultados inconclusos (Fig. 13). Si bien el 3.67% (1,998 loci) del total de secuencias agrupadas en el ensamble *Denovo\_97* fue asignado a este grupo, la diversidad de familias vegetales encontradas, dificulta la formulación de hipótesis sobre la presencia de DNA en el sistema digestivo de los escarabajos debido a su dieta herbívora. Es probable que algunas familias dominantes en el ensamble *Denovo\_97* como Fabaceae, Malvaceae o Solanaceae se puedan relacionar con la sobrerrepresentación que existe de estos taxa en la base de datos del NCBI, principalmente porque estas familias contienen representantes de importancia económica y agrícola altamente estudiados.

Las especies con mayor concentración de secuencias homólogas provenientes de los 1,998 loci de Streptophyta recuperados por *Denovo\_97*, fueron *D. micans* con 905 secuencias homólogas, *D. valens* con 703 y *D. frontalis* con 505 (Fig. 15). Entre las familias encontradas por este ensamble, la más numerosa fue Fabaceae con 442 loci, mientras que en Pinaceae, típicamente infectada por estos escarabajos, se encontraron únicamente 34 loci (Fig. 13). Entre estos 34 loci correspondientes a Pinaceae, se encontraron 124 secuencias homólogas del género *Pinus* con más de 97% de similitud y repartidas entre seis especies de *Dendroctonus* (*D. terebrans*, *D. murrayanae*, *D. micans*, *D. valens*, *D. rufipennis* y *D. approximatus* por orden de abundancia) así como 45 secuencias homólogas del género *Picea* presentes en seis especies (*D. micans*, *D. terebrans*, *D. rufipennis*, *D. punctatus*, *D. valens* y *D. approximatus*). Estos dos géneros, al igual que *Solanum*, destacan por su abundancia en comparación con los otros 14 géneros de Streptophyta con un 97% de similitud que solo obtuvieron menos de una decena de secuencias homólogas (p ej: *Vitis*, *Cucumis*, *Arachis*, *Pyricularia*, *Anticlea* y otros presentes en el Anexo 6).

De acuerdo con la literatura, *D. terebrans*, *D. valens* y *D. approximatus* infectan diferentes especies del género *Pinus*, mientras que *D. micans*, *D. murrayanae*, *D. punctatus* y *D. ruffipennis* son especies que pertenecen al mismo clado y que sólo pueden infectar al género *Picea*. Es posible que se encontraran loci de ambos géneros de pináceas repartidos entre especies que infectan sólo a uno de ellos debido a la proximidad filogenética de los árboles infectados y a la resolución de la determinación taxonómica a partir de RADSeq (Germandt *et al.*, 2008).

A pesar de que se ha logrado recuperar exitosamente DNA ambiental proveniente de interacciones entre plantas e insectos como eventos de polinización rastreados a partir de muestras de

miel (Hawkins et al., 2015; De Vere et al., 2017), el estudio de DNA vegetal asociado a muestras fecales de animales se ha restringido únicamente en grupos de talla grande debido a las dificultades de muestreo (Banerjee *et al.*, 2021). También se deben tomar en cuenta otros factores como la naturaleza del tejido vegetal del que proviene el DNA, ya que esto puede jugar un papel importante en la recuperación de secuencias. En el caso de las especies de escarabajo que se alimentan parcialmente de corteza externa, puede reducirse la posibilidad de identificar DNA del árbol infectado, además de que estos tejidos pueden contener metabolitos secundarios no digeridos por el insecto que se unen al DNA vegetal e imposibilitan su secuenciación (Echevarría-Machado *et al.*, 2005). A pesar de que la nutrición de las etapas larvales más tempranas de todas las especies del género depende del floema, algunas especies de *Dendroctonus* se caracterizan por desarrollarse de manera casi independiente a este tejido vegetal, tales como *D. brevicomis*, la cual sólo se alimenta brevemente del floema antes de mudarse a la corteza, y *D. approximatus*, cuyas larvas se desarrollan completamente a partir de corteza externa. Otras especies del grupo filogenético V como *D. adjunctus*, *D. mexicanus* y *D. frontalis* también se alimentan parcialmente de la corteza externa durante la etapa adulta (Six *et al.*, 2019).

Al analizar de manera general los sondeos realizados para el phylum Streptophyta a partir de los ensamblados *Denovo* y *DRef*, es lógico cuestionarse el origen de más de 70 familias que no parecen mantener una relación biológica más directa que la familia de árboles infectados, Pinaceae. Estos resultados confusos pueden atribuirse nuevamente al problema universal derivado del estado de las bases de datos nucleotídicas o bien, relacionarse con el tamaño reducido de los loci que se recuperan por medio de RADSeq. Se ha demostrado que este puede jugar un papel crucial en la identificación de grupos multicelulares, en particular aquellos de genomas complejos, de gran tamaño y con gran cantidad de elementos altamente repetitivos como los de las plantas (Pearman *et al.*, 2020; Wommack *et al.*, 2008).

## **Bacteria**

El Dominio Bacteria, caracterizado por contener una diversidad taxonómica y metabólica difícil de estimar (Locey y Lennon, 2016) fue mayormente representado en este estudio por phyla como Proteobacteria, Firmicutes, Actinobacteria y Tenericutes, en orden de abundancia recuperada por *Denovo\_97*. La dominancia del phylum Proteobacteria, particularmente a nivel de género,



concuerta con los estándares de diversidad en microbioma de insectos y con los estudios de microbioma central de este género (Kajtoch y Kotáskova, 2018; Hernández-García, 2017).

Entre las familias de Proteobacteria sondeadas en *Denovo\_97*, la más numerosa fue Pseudomonaceae, la cual ha sido reportada como parte del microbioma central del género *Dendroctonus* y fue corroborada al encontrarse en 14 de las 17 especies de este estudio (destacando a *D. micans*, *D. mexicanus* y *D. rufipennis* como la más abundante y exceptuando únicamente a *D. simplex*, *D. punctatus* y *D. pseudotsugae*) (Fig 17). Dentro de esta familia se han encontrado linajes aislados que son capaces de degradar celulosa, lípidos, enlaces éster, almidones y xilenos (Briones-Romero *et al.*, 2017).

La segunda familia de proteobacterias más numerosa fue Yersiniaceae, distribuida en todo el género pero mayormente concentrada en *D. approximatus*, *D. armandi* y *D. brevicomis*. Múltiples integrantes de esta familia han sido reportados como componentes principales del ectomicrobioma, particularmente de la comunidad bacteriana que habita en la cutícula de especies como *D. simplex* (Durand *et al.*, 2015) y del bacterioma intestinal de *D. frontalis* (Vasanthakumar *et al.*, 2006). Cabe destacar que algunas especies de esta familia han sido utilizadas como modelos en el estudio de la transmisión horizontal de simbiontes en artrópodos, por lo que su presencia a lo largo del género puede verse justificada por este fenómeno y por la falta de estudios de las comunidades microbianas de algunas especies de *Dendroctonus* (McAnally, 2016).

También se identificó a Moraxellaceae como una de las familias más abundantes, particularmente concentrada en *D. frontalis* con 1,059 secuencias homólogas, seguida por *D. micans* con sólo 17 y por el resto del género. Esta familia ha sido estudiada en algunas especies de la familia Curculionidae por su papel como degradador de azúcares complejos (Setiawan *et al.*, 2017), además se han encontrado ejemplares de esta familia, como *Acinetobacter* en *D. adjunctus*, *D. approximatus*, *D. brevicomis*, *D. frontalis* y *D. mexicanus*, y *Moraxella*, anteriormente encontrada en *D. mexicanus* (Hernández-García *et al.*, 2017). El género *Acinetobacter* se encontró entre los loci con 97% de similitud, particularmente debido a 628 secuencias homólogas concentradas en las muestras de *D. frontalis* (Fig. 18, Anexo 6).

Otras familias destacadas por su abundancia de loci, tales como Anaplasmataceae (533), Eriwiniaceae (491) y Enterobacteriaceae (267) han sido previamente reportadas como integrantes de importancia funcional en este género y en otras comunidades microbianas huéspedes de escarabajos de la subfamilia Scolytinae. A pesar de que de la familia Anaplasmataceae se identificó al género de mayor abundancia, *Wolbachia*, no se observó el mismo fenómeno con Eriwiniaceae y Enterobacteriaceae. Es posible que esto se deba a la diversidad de loci que no llegan a ser clasificables

a nivel de especie (con más de 97% de similitud) o a la subestimación que se ha sugerido que puede tener esta última familia en la microbiota intestinal de escarabajos al no ser identificada correctamente con los métodos de estudio disponibles (Ceja-Navarro *et al.*, 2015).

Se encontraron todos los géneros representantes del microbioma central de *Dendroctonus* entre las secuencias consenso con más de 97% de similitud recuperados por *Denovo\_97*, enlistados de mayor a menor por su abundancia: *Pseudomonas*, *Rahnella*, *Serratia*, *Pantoea*, *Raoultella* y *Enterobacter*. Algunos otros géneros de proteobacterias encontrados por este ensayo pertenecientes al microbioma central relajado incluyen *Erwinia*, *Acinetobacter* y a representantes de la familia *Enterobacteriaceae* (Hernández-García *et al.*, 2018). Entre algunas características metabólicas que se han encontrado por medio de cultivos aislados de estos simbiositos destacan *Pseudomonas* y *Raoultella* por su capacidad de fijar nitrógeno atmosférico (Morales-Jimenez *et al.*, 2009; Morales-Jimenez *et al.*, 2011), *Serratia*, *Rahnella* y *Pseudomonas* por degradar compuestos de defensa relacionados con las pináceas hospedadas del escarabajo (Xu *et al.*, 2016), *Acinetobacter* por su actividad lipolítica y esterasa y *Erwinia* por su papel en la síntesis de la feromona de agregación “verbenona” (Xu *et al.*, 2015).

Otro de los géneros más destacados por su abundancia en el ensayo fue *Wolbachia*, el cual ha sido profundamente estudiado por su relación ecológica como endosimbionte cosmopolita de otros artrópodos y nemátodos hospedados. De acuerdo a análisis estadísticos, al menos 20% de las especies de insectos existentes a nivel global son susceptibles a ser infectadas por este género de  $\alpha$ -proteobacterias (Hilgenboecker *et al.*, 2008). Además, sólo en el 0.15% de las especies de escarabajos conocidas se han hecho estudios para registrar la infección de *Wolbachia*, destacando a las familias Curculionidae y Chrysomelidae como las más estudiadas y que concentran un mayor número de infecciones reportadas (Kajtoch y Kotáskova, 2018). A pesar de que este género de bacterias ya había sido previamente reportado como simbiote mutualista o comensal en múltiples especies de *Dendroctonus*, la posibilidad de identificar loci con potencial patógeno a partir de datos altamente accesibles como los de secuenciación de genoma reducido, representa un área de oportunidad para arrancar investigaciones sobre el manejo de especies silvestres problemáticas poco estudiadas. La gran mayoría de las secuencias homólogas de este *Wolbachia* (1,734) se concentraron en *D. pseudotsugae*, con sólo menos del 15% asociados a *D. ponderosae* y *D. punctatus*. Esta bacteria había sido reportada únicamente en *D. ponderosae* (Lachowsky *et al.*, 2015), por lo que de confirmar su presencia en las otras dos especies huéspedes, este trabajo demostraría su potencial para proponer nuevas búsquedas y registros.

Al comparar el patrón de agrupación en la distribución de loci de las familias de Proteobacteria del ensamble *Denovo\_97*, con las relaciones filogenéticas reportadas por Godefroid y colaboradores en 2019, se encontró que pocas especies se agruparon con sus grupos filogenéticos, por ejemplo *D. rhizophagus*, *D. terebrans* y *D. valens* del grupo VI a nivel de phyla (Fig 15) y estas mismas a nivel de familia (Fig 17). Estos resultados coinciden con el sondeo de las comunidades bacterianas estudiadas a lo largo del género mediante secuenciación de amplicones de 16 rRNA, los cuales concluyeron que la mayor parte de la microbiota bacteriana se encuentra distribuida de manera general a lo largo de las especies consideradas, sin reflejar eventos claros de coevolución (Hernández-García *et al.*, 2017).

Además de los patrones de asociación por los que se organizaron a las especies de *Dendroctonus*, en los mapas de calor con diferentes dendrogramas (Fig 15, 17 y 18) se puede ver reflejada la similitud en las condiciones ecológicas que comparten estas poblaciones del género en lugar de una historia co-evolutiva en común, por ejemplo, la especie de árbol hospedero en el caso del nodo formado a partir de las familias de proteobacterias (Yersianaceae) compartidas entre *D. approximatus*, *D. brevicomis*, *D. ponderosae* y *D. armandii* (Fig 17). Si bien estas cuatro especies infectan a árboles del género *Pinus*, la dificultad de realizar estas abstracciones a partir de un set de datos previamente publicado radica en la ausencia de datos claves como la especie de árbol huésped, la altitud en la que se colectó cada muestra, la época del año o el estadio pre o post-emergente de estas mismas (Anexo 9) (Alberdi *et al.*, 2020).

Otro ejemplo de este problema radica en el rango altitudinal en el que habitan las especies que integran este mismo nodo: *D. approximatus* (1600 -3200 msnm), *D. brevicomis* (1100-3300 msnm), *D. ponderosae* (2400-2500 msnm) y *D. armandii* (1700-2500 msnm) se sobrelapan en un rango pequeño mayormente restringido por *D. ponderosae*, sin embargo al no contar con los datos exactos es difícil atribuir las similitudes en abundancia de proteobacterias a este factor de importancia ecológica (Armendáriz-Toledano *et al.*, 2018; Wang *et al.*, 2010). Habiendo tomado esto en cuenta, es posible sugerir que la naturaleza aleatoria de los datos trabajados y las similitudes entre la ecología de algunos de estas especies, son potencialmente responsables de la distribución de los loci de interés entre las especies de *Dendroctonus*.

Al analizar el total de categorías de familia de Proteobacteria identificadas por este método, es muy valioso destacar la abundancia de las categorías *no family* y *multiple*, la cual supera incluso a la familia más abundante, Pseudomonaceae. Al indagar en estos resultados es posible relacionarlos con los principales problemas que se han identificado en los métodos de asignación taxonómica en bacterias. El rezago histórico por el que ha pasado este grupo debido a la dificultad de su aislamiento

por medio de cultivos y a su tendencia a divergir genéticamente rápido (Rosselló-Mora y Amann, 2015; Konstantinidis y Roselló-Mora., 2015) puede parecer la explicación más directa de la prevalencia de estas categorías de familia, sin embargo, otros factores como el estado de curación y el sistema taxonómico empleado en las bases de datos taxonómicas también tienen una influencia importante (Cochrane *et al.*, 2016; Taylor, 2011). Hoy en día se han propuesto filogenias novedosas que desafían sustancialmente a la taxonomía en los dominios procariontes gracias a las aportaciones de las tecnologías de secuenciación de nueva generación durante las últimas décadas (Parks *et al.*, 2018; Rinke *et al.*, 2021). Estos hallazgos en materia de sistemática microbiana suelen aplicarse a bases de datos especializadas, por lo que el uso de bases de datos exhaustivas como la del NCBI o la del EMBL para fines metagenómicos enfocados se mantiene limitado. El emplear otros clasificadores alternos asociados a bases de datos microbianas puede traer consigo una clasificación más precisa, especialmente para estudios que busquen alcanzar una resolución a nivel de cepas o evitar la abundancia de etiquetas como *no taxa* o *multiple* (Cochrane *et al.*, 2016; Konstantinidis y Rosselló-Mora., 2015).

A partir de estas complicaciones con la determinación taxonómica se ha mantenido al gen ribosomal 16S como una herramienta efectiva y accesible, sin embargo para realizar la clasificación a nivel de especie y de cepa a partir de este código de barras en algunas ocasiones se requiere de la integración completa de regiones variables claves de este mismo gen y del desarrollo de algoritmos computacionales efectivos (Fiannaca *et al.*, 2018). Cuando se emplean secuencias metagenómicas “shotgun”, la clasificación se ve mejorada al incluir regiones más disímiles de las comunidades bacterianas, es decir, más fáciles de discriminar (Jovel *et al.*, 2016). Se ha propuesto en este caso, al emplear lecturas provenientes de diferentes loci, realizar un ensamble de la mayor cantidad de lecturas en fragmentos más grandes o contigs puede resultar muy beneficioso para su clasificación, incluso si estas lecturas provienen de diferentes muestras (Jovel *et al.*, 2016), tal como se ha realizado en este trabajo con RADSeq.

Algunos otros phyla bacterianos destacados por su relación con la microbiota del género *Dendroctonus* incluyen a Firmicutes con 196 loci, y Actinobacteria con 45. Algunos géneros pertenecientes a los phyla de bacterias Firmicutes (*Paenibacillus*, *Clostridium*, *Anaerococcus*, *Microbacterium*, y *Exiguobacterium*) y Actinobacteria (*Propionibacterium*, *Curtobacterium*, *Turicella*, *Corinebacterium*, y *Micrococcus*) han sido previamente reportados como los grupos más abundantes de la microbiota a lo largo del género *Dendroctonus*, sólo superados por Proteobacteria (Hernandez-García *et al.*, 2017), sin embargo no fue posible identificar loci a nivel de género pertenecientes a estos phyla por medio de este sondeo.

Por otra parte, se encontraron 208 loci de Apicomplexa en este estudio, los cuales pueden relacionarse con géneros de neograrinas parásitas como: *Matesia*, reportada previamente en *D. micans*, o bien con otros géneros de parásitos presentes en otras especies de escarabajos descortezadores (Yaman y Radek., 2015).

## **Nematoda**

Uno de los grupos que frecuentemente es descartado de los estudios de la microbiota, posiblemente debido a la dominancia de estudios de *metabarcoding* de bacterias, es el phylum Nematoda. Estos organismos han sido señalados como un componente esencial para la comprensión de las relaciones ecológicas que mantienen múltiples poblaciones de *Dendroctonus*, e incluso han sido propuestos como una posible área de oportunidad para el diseño de estrategias de control biológico debido a su papel parasítico (Mills, 1983).

Cuando se analiza el patrón de asociación de los loci de Nematoda entre las diferentes especies (Fig. 15), *D. adjunctus* resalta como la especie con una mayor abundancia. Este resultado puede ser relacionado con estudios en los que se han encontrado hasta 20 especies de nemátodos en un sólo ejemplar de esta especie de escarabajo (Massey, 1966), sin embargo, en múltiples integrantes de los grupos filogenéticos III, IV y V se ha reportado una relación frecuente con este phylum (Hofstatter, 2015).

Entre los 463 loci de nemátodos encontradas por el ensamble *Denovo\_97*, la familia más abundante fue Rhabditidae, la cual alberga muchas especies de endocomensales de artrópodos. Esta familia también contiene al género de nemátodos (>97% de similitud) con un mayor número de coincidencias: *Caenorhabditis* (Fig 17), cuya sobrerrepresentación en la base de datos nucleotídica del NCBI con respecto al resto del especies de este phylum puede estar íntimamente relacionada con este resultado. Es probable que el material genético de algunos de los simbiontes de la familia Rhabditidae que han sido previamente reportados en *Dendroctonus* (Carta *et al.*, 2010; Thong y Webster, 1983) se alineara con mayor facilidad en contra de las secuencias de referencia de *C. elegans* disponibles en la base de datos.

Además se encontraron otros dos géneros de nemátodos débilmente representados entre las diferentes especies de *Dendroctonus*: *Ditylenchus* y *Trichinella* con dos y un loci respectivamente. Se ha reportado que ejemplares de *Ditylenchus* pueden ser depredadores de las larvas de *D. valens* y *D. frontalis*, dos de las seis especies en las que se encontraron secuencias homólogas de este género, por

lo que se asume que estos nemátodos pueden ser encontrados en el mismo hábitat de diferentes poblaciones de *Dendroctonus* (Rose, 1996).

## **Perspectivas**

El presente estudio describe una aplicación de herramientas bioinformáticas generalizadas para el aprovechamiento de datos de RADSeq previamente generados, con el fin de realizar sondeos de la diversidad en la microbiota de insectos. El potencial de este método radica en la exploración de la microbiota de grupos para los que difícilmente existen secuencias genómicas disponibles o bien, datos de secuenciación de la microbiota generados por *metabarcoding* o secuenciación *shotgun*. La capacidad de tener este acercamiento previo a los grupos microbianos, o de interés biológico, puede ayudar a construir hipótesis sobre la distribución de la microbiota a lo largo de hospederos filogenéticamente relacionados (poblaciones o múltiples especies de un género) o bien, a realizar diseños experimentales enfocados en el aislamiento por medio de cultivos, o el estudio expandido de la microbiota, o el DNA ambiental asociado a hospederos poco estudiados.

Una de las principales pruebas que requiere este método para soportar con mayor firmeza su potencial, es la confirmación de los resultados encontrados, es decir de los loci asignados a diferentes grupos, por medio de una técnica alternativa como PCR dirigida o secuenciación de amplicones a partir hospederos provenientes del mismo muestreo. Además de utilizarse como un enfoque metagenómico para acercarse a grupos de interés por medio de datos de RADSeq, es conveniente tomar en consideración los siguientes rubros.

También se considera la posibilidad de realizar la búsqueda por medio de BLAST loci en particular de algunos grupos de interés y buscar si corresponden con alguna región en particular que sirva, por ejemplo, para diseñar primers o sondas y hacer otro tipo de estudios.

La capacidad de realizar análisis cofilogenéticos a partir de un tipo de secuenciación tan accesible como el RADSeq destaca entre las etapas más relevantes de este trabajo, por lo que su optimización y la posterior confirmación de las hipótesis generadas por medio de este tipo de sondeo deben ser tomadas en cuenta al reproducirlo.

## **Recomendaciones del muestreo y la obtención de secuencias**

Para considerar a RADSeq como una secuenciación parcial del metagenoma, sería necesario reducir la variación en los diferentes niveles de complejidad implicados en el muestreo. Esto se refiere

a la localidad de origen, la edad de los hospederos, el tiempo en el que se colectan e incluso la cantidad de información genómica que se posee del hospedero para ocupar como filtro y quedarse sólo con la información de la comunidad microbiana simbiote. Esto puede pensarse como un *trade off* de la búsqueda de secuencias del microbioma en RADseq, ya que es posible ocupar set de datos provenientes de muestras heterogéneas como muestras de diferentes especies, localidades o de colecciones; sin embargo se debe tener cuidado al proponer su papel como integrantes del microbioma, ya que no se cuenta con la resolución necesaria para realizar una asignación taxonómica profunda ni con el control de las variables ambientales.

El estado de conservación de las muestras juega un papel crucial en este método, ya que se ha reportado que al emplear tejidos de vertebrados, las muestras deben contener al menos 40,000 secuencias identificables para empezar a rastrear loci metagenómicos (Holmes y Rabosky, 2018), lo cual puede ser una limitante al trabajar con muestras provenientes de museo o de condiciones poco óptimas. Al realizar un diseño experimental para la exploración de secuencias de RADSeq, es importante replicar las prácticas de muestreo recomendadas para mantener un rigor estadístico, tales como contar con la mayor cantidad de ejemplares de la misma especie o población, procurar que los replicados provengan de un mismo sitio de colecta y siempre reportar la información completa sobre las condiciones de muestreo (Ballare *et al.*, 2018).

Partiendo de las dificultades que tiene trabajar con material fresco de una gran cantidad de grupos de interés ecológico, es importante tomar en cuenta diferentes métodos para procurar una mejor preservación del material genómico del insecto y de su microbiota. La discusión sobre el mejor método de preservación para la secuenciación del microbioma de insectos se ha mantenido abierta. Se ha reportado que el mejor método para obtener una representación mayor de comunidades bacterianas se obtiene al emplear ejemplares frescos, sin importar el empleo de técnicas de disección del sistema digestivo (Yang *et al.*, 2021; De Cock *et al.*, 2019). Al comparar los métodos de preservación de las muestras se han encontrado resultados diversos, se ha comparado la influencia de los métodos de conservación en la secuenciación de amplicones bacterianos de hemípteros cuyo DNA fue extraído del individuo completo, en particular en dos plataformas, Illumina y PacBio. En ambos casos se reporta una mejor representación de la comunidad microbiana al preservar los ejemplares en bromuro de hexadeciltrimetilamonio (CTAB) y en etanol al 99% que al ocupar otros métodos como el secado por aire o el medio RNALater (Yang *et al.*, 2021). Sin embargo también se ha encontrado que la preservación de moscas de la fruta (Diptera) en etanol al 70% a -20 °C entre 12 y 18 semanas puede tener también un efecto negativo en la secuenciación de amplicones. En particular se encontró una

abundancia menor de taxa dominantes y de taxa raros a comparación con la secuenciación a partir de ejemplares frescos (De Cock *et al.*, 2019).

Aún es necesario comparar el efecto que puede tener la secuenciación de tejidos específicos en insectos en comparación a la de la digestión de los individuos completos, como se realizó en este trabajo, sin embargo de acuerdo a los resultados del trabajo de Holmes y Rabosky del 2018, la secuenciación de RADSeq a partir de diferentes tejidos de reptiles no parece tener un efecto en la cantidad de loci microbianos recuperados.

Por último, es claro que las condiciones del ensayo como la construcción de la librería genómica y la plataforma de secuenciación elegida tienen un impacto significativo dependiendo de los objetivos del sondeo y la resolución taxonómica que este aspire alcanzar. Aunque el estudio de la microbiota a partir de este tipo de datos representa un reto y es posible también seleccionar el tamaño de las lecturas ensambladas por ipyrad al definir los parámetros de los ensambles, por lo que se sugiere que estas variables deben ser evaluadas para diseñar futuros sondeos.

## **Recomendaciones técnicas**

Una de las principales observaciones que deben tomarse en cuenta al implementar este método para sondear la diversidad microbiana asociada a datos de RADSeq, es la cantidad de recursos computacionales que se le asigna a la segunda parte de este método, en lo particular a los análisis de BLAST y en menor medida, a la asignación taxonómica realizada en R con el paquete *taxize*. Aunque es conveniente realizar análisis de optimización computacional para los procesos más exigentes que propone este método, a continuación se presentan algunas recomendaciones sobre este tema.

Los resultados preliminares a partir de los cuales se evaluaron múltiples tipos de ensamble fueron realizados con sólo 2 de los cores disponibles en el clúster de supercómputo de CONABIO que fue proporcionado para trabajar inicialmente este proyecto. No obstante, al realizar nuevamente estas pruebas con al menos 16 cores en *Miztli*, el clúster de supercómputo de la UNAM, se encontraron resultados extremadamente favorables con los ensambles *Denovo* y *DRef*. Es importante contrastar que en los ensambles de *GRef* no se observó cambio alguno al asignarle un número diferente de cores a la prueba de BLAST, por lo que el potencial de ensamblar los datos empleando como referencia secuencias del genoma completo de simbioses previamente reportados puede ser muy alto cuando los recursos técnicos son limitados.



Algunos estudios de optimización computacional asociados al uso de BLAST en clusters de supercómputo, han encontrado un desempeño desfavorable mediante la paralelización del trabajo estándar que ofrece el programa. Este proceso está limitado únicamente entre múltiples cores de una sola computadora (nodo), por lo que se propone la integración de softwares adicionales llamados “wrappers” diseñados para la resolución de esta limitante y la aceleración del proceso al repartir las secuencias y la base de datos entre los múltiples nodos que pueden integrar un clúster de supercómputo (Yim y Cushman, 2017). La optimización de este proceso cobra aún mayor importancia al emplear archivos multi-fasta de gran tamaño como los que se generan por medio de algunas de las estrategias de ensamble propuestas (*de novo, de novo -ref*), las cuales pueden no generar un output completo de no contar con los recursos computacionales necesarios.

De igual forma se considera crucial el asignar suficientes de recursos computacionales a la segunda parte del script publicado por Holmes en el 2018, con la cual se recuperan los ID's generados por medio de la prueba de BLAST para realizar una asignación taxonómica en todos los niveles de cada loci con un porcentaje de similitud de al menos 90%. Aunque esta parte del trabajo se realizó en una computadora personal de capacidad promedio (8 GB de RAM y procesador de 4 núcleos), es importante permitir que este se complete sin interrupciones o tareas paralelas debido a la sensibilidad que este programa tiene al realizar la clasificación de los ID's y a la posibilidad de interpretar un output incompleto como resultado debido a la dificultad de analizar y comparar archivos de tamaño masivo.

Aunque el desarrollo de este método está sujeto a determinados recursos tecnológicos, el presente trabajo deja claro su potencial exploratorio y sienta las bases para expandir el tipo de inferencias o hipótesis biológicas que se pueden rescatar del mismo. Durante los últimos años, la cantidad de datos bioinformáticos generados a nivel global ha crecido desmedidamente, por lo que parece oportuno promover su aprovechamiento y su relación con problemas biológicos vigentes como el manejo de los escarabajos descortezadores.

## CONCLUSIONES

- Se analizaron secuencias de RADSeq asociadas a diecisiete especies del género *Dendroctonus* provenientes de muestreos independientes en Norteamérica y Eurasia. Por medio de diferentes métodos de ensamble de secuencias, se encontraron 9,704 loci considerados como informativos, ya sea por su posible papel como integrantes del microbioma o por su relación con el entorno biológico de estos escarabajos.

- Entre los diferentes métodos de ensamblaje propuestos, el ensamblaje *de novo* con un valor de CT alto (0.97) permite encontrar una mayor cantidad de loci y una mayor diversidad de taxa.
- Los ensamblajes *Denovo - referencia* tienen un mejor desempeño con un valor de de CT de 0.95, sin embargo este método se ve sujeto a la calidad de la secuencia genómica de referencia utilizada por lo que la posibilidad de realizar estudios más enfocados a partir de un catálogo más amplio y mejor trabajado de secuencias de referencia se mantiene abierta.
- El realizar estos análisis a partir de una mayor cantidad de muestras parece tener un efecto más relevante en la recuperación de loci metagenómicos que la identidad específica de las muestras, como se ejemplifica con los ensamblajes *DPond*.
- El empleo de secuencias de referencia de genoma completo de simbiontes durante el ensamblaje de loci en la pipeline *ipyrad*, mediante la disminución del parámetro *mindepth\_makrule*, permite optimizar la búsqueda y enfocar en secuencias de baja profundidad.
- La determinación taxonómica puede verse influenciada por factores como el tamaño de los fragmentos de RADSeq con los que se trabaja y el grupo que se busca recuperar, por lo que algunos taxa procariontes pueden tener más dificultades para alcanzar una resolución taxonómica precisa debido a los sesgos de las bases de datos empleadas o a la dificultad de caracterizar estos ejemplares no cultivables y con altas tasas de mutación; mientras que los eucariotas pueden presentar errores hasta estos niveles debido a la presencia de elementos altamente repetitivos en sus genomas.
- Se encontraron representantes de más de 50 phyla de importancia biológica para el género *Dendroctonus* como Proteobacteria, Firmicutes, Actinobacteria, Tenericutes, Ascomycota, Basidiomycota, Nematoda y Streptophyta.
- Se encontraron todos los integrantes del microbioma central estricto previamente reportado para trece especies de *Dendroctonus* representados en las diecisiete especies estudiadas: *Enterobacter*, *Pantoea*, *Pseudomonas*, *Rahnella*, *Raoultella* y *Serratia*, además de

algunos representantes del microbioma central relajado como *Erwinia*, *Acinetobacter* y representantes de la familia Enterobacteriaceae.

- Se encontraron loci representantes del género *Wolbachia* en *D. pseudotsugae* y *D. punctatus*, especies reportadas donde no habían sido previamente reportadas. Además se reportó nuevamente su presencia en *D. ponderosae*.
- No se encontraron patrones de asociación de los loci microbianos entre las diferentes especies de *Dendroctonus* que sugieran eventos de coespeciación de grupos filogenéticos completos, sin embargo, se ubicaron loci asociados a los árboles hospederos como *Pinus* y *Picea* provenientes de especies que se caracterizan por infectar a estos géneros. Es posible que algunas especies agrupadas por su patrón de asociación con loci microbianos compartan factores ecológicos como la especie de árbol que infectan o la altura en la que habitan.
- Ante el incesante desarrollo de las diferentes técnicas de secuenciación masiva en paralelo, la secuenciación de RADSeq continúa siendo una de las opciones más accesibles y flexibles con respecto al diseño del muestreo. Gracias a estas ventajas y a la posibilidad de ocupar secuencias previamente generadas para realizar sondeos del microbioma, se propone que este método puede ser un parteaguas en este tipo de estudios para especies silvestres con el cual se podrán formular preguntas más precisas con herramientas de mayor resolución metagenómica.

## REFERENCIAS

- Adams, A. *et al.* (2009) Effects of symbiotic bacteria and tree chemistry on the growth and reproduction of bark beetle fungal symbionts. *Canadian Journal of Forest Research*. 39(6): 1133-1147. <https://doi.org/10.1139/X09-034>
- Adams, A. S., Boone, C. K., Bohlmann, J., & Raffa, K. F. (2011). Responses of bark beetle-associated bacteria to host monoterpenes and their relationship to insect life histories. *Journal of chemical ecology*, 37(8), 808-817.
- Adams, A., Aylward, F., Adams, S., Erbilgin, N., Aukema, B., Currie, C, ... & Raffa, K.. (2013). Mountain pine beetles colonizing historical and naive host trees are associated with a bacterial community highly enriched in genes contributing to terpene metabolism. *Applied and environmental microbiology*, 79(11), 3468-3475.

- Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990) "Basic local alignment search tool." *J. Mol. Biol.* 215:403-410. PubMed
- Alberdi, A., Andersen, S. B., Limborg, M. T., Dunn, R. R., & Gilbert, M. T. P. (2021). Disentangling host–microbiota complexity through hologenomics. *Nature Reviews Genetics*, 1-17.
- Andrews, K., Good, J., Miller, M., Luikart, G., & Hohenlohe, P. (2016). Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics*, 17(2), 81–92. doi:10.1038/nrg.2015.28 }
- Armendáriz-Toledano, F., Zúniga, G., García- Román, L. J., Valerio-Mendoza, O. y Navarrete, P.G. (2018) *Guía ilustrada para identificar a las especies del género Dendroctonus presentes en México y Centroamérica*. Instituto Politécnico Nacional. CDMX, México.
- Aunen, D. K. & Hoekstra, R.F. (2007) The evolution of obligate mutualism: if you can beat'em, join'em. *Trends in Ecology and Evolution*. 22, 506 - 509.
- Aschenbrenner, I. A., Cernava, T., Berg, G., & Grube, M. (2016). *Understanding Microbial Multi-Species Symbioses*. *Frontiers in Microbiology*, 7. doi:10.3389/fmicb.2016.00180
- Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., ... & Johnson, E. A. (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PloS one*, 3(10), e3376.
- Banerjee, P., Stewart, K. A., Antognazza, C. M., Bunholi, I. V., Deiner, K., Barnes, M. A., ... & Chen, C. Y. (2021). Plant-animal interactions in the era of environmental DNA (eDNA)—a review. *Authorea Preprints*.
- Balvočiūtė, M., & Huson, D. H. (2017). SILVA, RDP, Greengenes, NCBI and OTT—how do these taxonomies compare?. *BMC genomics*, 18(2), 1-8
- Bentz, B. J., yZ Six, D. L. (2006). Ergosterol content of fungi associated with *Dendroctonus ponderosae* and *Dendroctonus rufipennis* (Coleoptera: Curculionidae, Scolytinae). *Annals of the Entomological Society of America*, 99(2), 189-194.
- Bentz, B. J. (2008). Bark Beetles in the Genus *Dendroctonus*. En J. L. Capinera (Ed.), *Encyclopedia of Entomology* ( 379–381). Springer.
- Bleiker, K. & Six, D.L. (2007) Dietary benefits of fungal associates to an eruptive herbivore: potential implications of multiple associates on host population dynamics. *Environmental Entomology*, 36, 1384–1396.
- Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30 (15), 2114-2120.
- Boone, C. K., Six, D. L., Zheng, Y., & Raffa, K. F. (2014). Parasitoids and dipteran predators exploit volatiles from microbial symbionts to locate bark beetles. *Environmental Entomology*, 37(1), 150-161.
- Breitwieser, F. P., Lu, J., & Salzberg, S. L. (2019). A review of methods and databases for metagenomic classification and assembly. *Briefings in bioinformatics*, 20(4), 1125-1136.

Bridges, J. R., Marler, J. E., & McSparrin, B. H. (1984). A quantitative study of the yeasts and bacteria associated with the laboratory-reared *Dendroctonus frontalis* Zimm.(Coleopt., Scolytidae) 1. *Zeitschrift für angewandte Entomologie*, 97(1-5), 261-267.

Briones-Roblero, C. I., Hernández-García, J. A., Gonzalez-Escobedo, R., Soto-Robles, L. V., Rivera-Orduña, F. N., & Zúñiga, G. (2017). Structure and dynamics of the gut bacterial microbiota of the bark beetle, *Dendroctonus rhizophagus* (Curculionidae: Scolytinae) across their life stages. *PLOS ONE*, 12(4), e0175470. doi:10.1371/journal.pone.0175470

Bridges, J. R. (1983). Mycangial Fungi of *Dendroctonus frontalis* (Coleoptera: Scolytidae) and their Relationship to Beetle Population Trends. *Environmental Entomology*, 12(3), 858-861.

Briones-Roblero, C. I., Rodríguez-Díaz, R., Santiago-Cruz, J. A., Zúñiga, G., & Rivera-Orduña, F. N. (2017). Degradation capacities of bacteria and yeasts isolated from the gut of *Dendroctonus rhizophagus* (Curculionidae: Scolytinae). *Folia microbiologica*, 62(1), 1-9.

Briones-Roblero, C. I., Hernández-García, J. A., Gonzalez-Escobedo, R., Soto-Robles, L. V., Rivera-Orduña, F. N., & Zúñiga, G. (2017). Structure and dynamics of the gut bacterial microbiota of the bark beetle, *Dendroctonus rhizophagus* (Curculionidae: Scolytinae) across their life stages. *PLoS one*, 12(4), e0175470.

Brown, J., Pirrung, M., & McCue, L. A. (2017). FQC Dashboard: integrates FastQC results into a web-based, interactive, and extensible FASTQ quality control tool. *Bioinformatics*, 33(19), 3137-3139.

Brusca, R. C., Brusca, G.J. y Shuster, S.M. (2016) "*Blastocolemoates and Other Phyla*" en *Invertebrates*. 338-340 pp. Ed. McGrawHill. EEUU.

Bin, L., Lin, W., Muzammil, H., Xiaoling, Z., Jianqing, T., Sha, Q., ... & Meichun, X. (2019). Microbiota analysis revealed vertical transmission and microbial adjusting function of symbiotic fungus in the cochranattelabid weevil fungiculture. *Science China Life Sciences*, 62(12), 1717-1721.

Cardoza, Y. J., Moser, J. C., Klepzig, K. D., & Raffa, K. F. (2008). Multipartite symbioses among fungi, mites, nematodes, and the spruce beetle, *Dendroctonus rufipennis*. *Environmental Entomology*, 37(4), 956-963.

Carlton, J. M., Angiuoli, S. V., Suh, B. B., Kooij, T. W., Pertea, M., Silva, J. C., ... & Carucci, D. J. (2002). Genome sequence and comparative analysis of the model rodent malaria parasite *Plasmodium yoelii yoelii*. *Nature*, 419(6906), 512-519.

Castello, J.. (1976) Isolation of *Cryptoporus volvatus* and *Fomes pinicola* from *Dendroctonus pseudotsugae*. *Phytopathology* 66 (12)1431-1434

Chamberlain, S., Szoecs, E., Foster, Z., Arendsee, Z., Boettiger, C., Ram, K., Bartomeus, I., Baumgartner, J., O'Donnell, J., Oksanen, J., Greshake-Tzovaras, B., Marchand, P., Tran, V., Salmon, M., Li, G., y Grenié, M. (2020) taxize: Taxonomic information from around the web. R package version 0.9.98. <https://github.com/ropensci/taxize>

Cole, J. R., Chai, B., Farris, R. J., Wang, Q., Kulam, S. A., McGarrell, D. M., ... & Tiedje, J. M. (2005). The Ribosomal Database Project (RDP-II): sequences and tools for high-throughput rRNA analysis. *Nucleic acids research*, 33(suppl\_1), D294-D296.

- Coppedge, B. R., Stephen, F. M., & Felton, G. W. (1995). Variation in female southern pine beetle size and lipid content in relation to fungal associates. *The Canadian Entomologist*, 127(2), 145-154.
- Cochrane, G., Karsch-Mizrachi, I., Takagi, T., & Sequence Database Collaboration, I. N. (2016). The international nucleotide sequence database collaboration. *Nucleic acids research*, 44(D1), D48-D50.
- Craud, A., Gautier, M., Rossi, J. P., Rasplus, J. Y., & Gouzy, J. (2016). RADIS: analysis of RAD-seq data for interspecific phylogeny. *Bioinformatics*, 32(19), 3027-3028.
- Cuadros-Orellana S, Leite LR, Smith A, Medeiros JD, Badotti F, et al. (2013) Assessment of Fungal Diversity in the Environment using Metagenomics: a Decade in Review. *Fungal Genome Biol* 3(110). doi:10.4172/2165-8056.1000110
- Davis, T. S., & Hofstetter, R. W. (2011). Reciprocal interactions between the bark beetle-associated yeast *Ogataea pini* and host plant phytochemistry. *Mycologia*, 103(6), 1201-1207.
- De Cock, M., Virgilio, M., Vandamme, P., Augustinos, A., Bourtzis, K., Willems, A., & De Meyer, M. (2019). Impact of sample preservation and manipulation on insect gut microbiome profiling. A test case with fruit flies (Diptera, Tephritidae). *Frontiers in microbiology*, 10, 2833.
- Delalibera Jr, I., Handelsman, J. O., & Raffa, K. F. (2005). Contrasts in cellulolytic activities of gut microorganisms between the wood borer, *Saperda vestita* (Coleoptera: Cerambycidae), and the bark beetles, *Ips pini* and *Dendroctonus frontalis* (Coleoptera: Curculionidae). *Environmental Entomology*, 34(3), 541-547.
- del-Val, E. & Sáenz-Romero, C. (2017) Insectos descortezadores (Coleoptera: Curculionidae) y cambio climático: problemática actual y perspectivas en los bosques templados. *Revista Especializada en Ciencias Químico-Biológicas*, 20(2): 53-60, UNAM.
- DeSantis, T. Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E. L., Keller, K., ... & Andersen, G. L. (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Applied and environmental microbiology*, 72(7), 5069-5072.
- DiGuistini, S., et al (2011). Genome and transcriptome analyses of the mountain pine beetle-fungal symbiont *Grosmannia clavigera*, a lodgepole pine pathogen. *Proceedings of the National Academy of Sciences*, 108(6), 2504–2509. doi:10.1073/pnas.1011289108
- Donovan, P. D., Gonzalez, G., Higgins, D. G., Butler, G., & Ito, K. (2018). Identification of fungi in shotgun metagenomics datasets. *PLoS ONE*, 13(2), 1–16.
- Douglas, A. E. (2009). The microbial dimension in insect nutritional ecology. *Functional Ecology*, 23(1), 38-47.
- Durand, A. A., Bergeron, A., Constant, P., Buffet, J. P., Déziel, E., & Guertin, C. (2015). Surveying the endomicrobiome and ectomicrobiome of bark beetles: The case of *Dendroctonus simplex*. *Scientific reports*, 5(1), 1-11
- Durand, A. A. (2018). *Caractérisation du microbiome (bactéries et champignons) associé au dendroctone du mélèze, Dendroctonus simplex Leconte* (Doctoral dissertation, Université du Québec, Institut national de la recherche scientifique).

- Eaton, D. A. R. (2014). PyRAD: Assembly of de novo RADseq loci for phylogenetic analyses. *Bioinformatics*, 30(13), 1844–1849.
- Eaton DAR & Overcast I. (2020) "ipyrad: Interactive assembly and analysis of RADseq datasets." *Bioinformatics* 36, (8) 2592–2594, <https://doi.org/10.1093/bioinformatics/btz966>
- Echevarria-Machado, I., Sanchez-Cach, L.A., Hernandez-Zepeda, C., Rivera-Madrid, R., & MorenoValenzuela, O.A. (2005). A simple and efficient method for isolation of DNA in high mucilaginous plant tissues. *Molecular Biotechnology*, 31(2), 129-135. <https://doi.org/10.1385/MB:31:2:129>.
- Eisen, J. A. (2007). "Environmental shotgun sequencing: its potential and challenges for studying the hidden world of microbes". *PLOS Biology*. 5(3): e82. doi:10.1371/journal.pbio.0050082. PMC 1821061. PMID 17355177
- Etter, P. D., Preston, J. L., Bassham, S., Cresko, W. A., & Johnson, E. A. (2011). Local de novo assembly of RAD paired-end contigs using short sequencing reads. *PloS one*, 6(4), e18561.
- Escobar-Zepeda, A., Vera-Ponce de León, A., & Sanchez-Flores, A. (2015). The road to metagenomics: from microbiology to DNA sequencing technologies and bioinformatics. *Frontiers in genetics*, 6, 348.
- Farrell, B. D., Sequeira, A. S., O'Meara, B. C., Normark, B. B., Chung, J. H., & Jordal, B. H. (2001). The evolution of agriculture in beetles (Curculionidae: Scolytinae and Platypodinae). *Evolution*, 55(10), 2011-2027.
- Fiannaca, A., La Paglia, L., La Rosa, M., Bosco, L., Renda, G., Rizzo, R., ... & Urso, A. (2018). Deep learning models for bacteria taxonomic classification of metagenomic data. *BMC bioinformatics*, 19(7), 61-76.
- Fraser, J. D. (2011). *Cold tolerance and seasonal gene expression in Dendroctonus ponderosae* (Doctoral dissertation, University of Northern British Columbia).
- Gernandt, D. S., Magallón, S., Geada López, G., Zerón Flores, O., Willyard, A., & Liston, A. (2008). Use of simultaneous analyses to guide fossil-based calibrations of Pinaceae phylogeny. *International Journal of Plant Sciences*, 169(8), 1086-1099.
- Gibson, C. M., & Hunter, M. S. (2010). Extraordinarily widespread and fantastically complex: comparative biology of endosymbiotic bacterial and fungal mutualists of insects. *Ecology letters*, 13(2), 223-234.
- Giblin-Davis, R. M., Kanzaki, N., & Davies, K. A. (2013). Nematodes that Ride Insects: Unforeseen Consequences of Arriving Species §. *Florida Entomologist*, 96(3), 770-780.
- Godefroid, M., Rasplus, J. Y., & Rossi, J. P. (2016). Is phylogeography helpful for invasive species risk assessment? The case study of the bark beetle genus *Dendroctonus*. *Ecography*, 39(12), 1197-1209.
- Godefroid, M., Meseguer, A. S., Saune, L., Genson, G., Streito, J. C., Rossi, J. P., ... & Rasplus, J. Y. (2019). Restriction-site associated DNA markers provide new insights into the evolutionary history of the bark beetle genus *Dendroctonus*. *Molecular phylogenetics and evolution*, 139, 106528.

- Gontier, N. (2016). Symbiosis, History of. In: Kliman, RL (ed) *Encyclopedia of Evolutionary Biology* (272-281). Elsevier
- Grégoire, J. (1998) Greater european spruce beetle, *Dendroctonus micans*, pp 445 - 478. En A.A. Berryman (Ed.), *Dynamics of insects population patterns, causes, implications*. Plenum Press, NY.
- Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., ... & MacManes, M. D. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature protocols*, 8(8), 1494.
- Harrington, T. C. (1993). Diseases of conifers caused by species of Ophiostoma and Leptographium. *Ceratocystis and Ophiostoma: taxonomy, ecology, and pathogenicity*, 161-172.
- Harrington, T. C. (2005). Ecology and evolution of mycophagous bark beetles and their fungal partners. En: Bega F. E. y Blackwell, M.(Eds.), *Ecological and Evolutionary Advances in Insect- Fungal Associations*,. Oxford University Press
- Hawksworth, D. L., & Lücking, R. (2017). Fungal diversity revisited: 2.2 to 3.8 million species. *Microbiology spectrum*, 5(4), 5-4.
- Hernández-García, J. A., Briones-Roblero, C. I., Rivera-Orduña, F. N., & Zúñiga, G. (2017). Revealing the gut bacteriome of Dendroctonus bark beetles (Curculionidae: Scolytinae): diversity, core members and co-evolutionary patterns. *Scientific Reports*, 7(1). doi:10.1038/s41598-017-14031-6
- Hernández-García, J. A., Gonzalez-Escobedo, R., Briones-Roblero, C. I., Cano-Ramírez, C., Rivera-Orduña, F. N., & Zúñiga, G. (2018). Gut bacterial communities of Dendroctonus valens and D. mexicanus (Curculionidae: Scolytinae): a metagenomic analysis across different geographical locations in Mexico. *International journal of molecular sciences*, 19(9), 2578
- Holmes, I. y Davis, A. (2018) Natural history bycatch: a pipeline for identifying metagenomic sequences in RADseq data. PeerJ DOI 10.7717/peerj.4662
- Hooper, R., Brealey, J. C., van der Valk, T., Alberdi, A., Durban, J. W., Fearnbach, H., ... Guschanski, K. (2018). Host-derived population genomics data provides insights into bacterial and diatom composition of the killer whale skin. *Molecular Ecology*. doi:10.1111/mec.14860
- Hofstetter, R. W., Dinkins-Bookwalter, J., Davis, T. S., & Klepzig, K. D. (2015). Symbiotic associations of bark beetles. En Vega, F. E., & Hofstetter, R. W. (Eds.), *Bark Beetles* (pp. 209-245). Academic Press.
- Hu, X., Li, M., & Chen, H. (2015). Community structure of gut fungi during different developmental stages of the Chinese white pine beetle (*Dendroctonus armandi*). *Scientific Reports*, 5(1). doi:10.1038/srep08411
- Hulcr, J., Rountree, N. R., Diamond, S. E., Stelinski, L. L., Fierer, N., & Dunn, R. R. (2012). Mycangia of Ambrosia Beetles Host Communities of Bacteria. *Microbial Ecology*, 64(3), 784-793. doi:10.1007/s00248-012-0055-5
- Huler, J. & Stelinski, L.L. (2017) The ambrosia symbiosis: from evolutionary ecology to practical management. *Annual Review of Entmology*. 62, 285 - 303



- Hunt, D. W. A., & Borden, J. H. (1990). Conversion of verbenols to verbenone by yeasts isolated from *Dendroctonus ponderosae* (Coleoptera: Scolytidae). *Journal of Chemical Ecology*, 16(4), 1385-1397.
- Jermiin, L. S., Olsen, G. J., Mengersen, K. L., & Easteal, S. (1997). Majority-rule consensus of phylogenetic trees obtained by maximum-likelihood analysis. *Molecular Biology and Evolution*, 14(12), 1296-1302.
- Jovel, J., Patterson, J., Wang, W., Hotte, N., O'Keefe, S., Mitchel, T., ... & Wong, G. K. S. (2016). Characterization of the gut microbiome using 16S or shotgun metagenomics. *Frontiers in microbiology*, 7, 459.
- Hou, Z., Dong, Y., Shi, F., Xu, Y., Ge, S., Tao, J., ... & Zong, S. (2021). Seasonal shifts in cold tolerance and the composition of the gut microbiome of *Dendroctonus valens* LeConte occur concurrently. *Forests*, 12(7), 888.
- Hurst, G. y Darby, A. (2011) The inerted microbiota of arthropods, and their importance in understanding resistance and immunity. En: Rolff, J.y Reynolds, S. (Eds.), *Insect Infection and Immunity: Evolution, Ecology, and Mechanisms*. (119-137). Oxford University Press.
- Keeling, C. I., Yuen, M. M., Liao, N. Y., Docking, T. R., Chan, S. K., Taylor, G. A., ... & Bohlmann, J. (2013). Draft genome of the mountain pine beetle, *Dendroctonus ponderosae* Hopkins, a major forest pest. *Genome biology*, 14(3), R27.
- Kinn, D. N., & Stephen, F. M. (1981). The incidence of endoparasitism of *Dendroctonus frontalis* Zimm.(Coleoptera: Scolytidae) by *Contortylenchus brevicomi* (Massey) Rühm (Nematoda: Sphaerulariidae) 1. *Zeitschrift für Angewandte Entomologie*, 91(1-5), 452-458.
- Klepzig, K. D., Moser, J. C., Lombardero, F. J., Hofstetter, R. W., & Ayres, M. P. (2001). Symbiosis and competition: complex interactions among beetles, fungi and mites. *Symbiosis*.
- Knell JD & Allen GE (1978) Morphology and ultrastructure of *Unikaryon minutum* sp.n. (Microsporidia: Protozoa) a parasite of southern pine beetle, *Dendroctonus frontalis*. *Acta Protozool* 17:271-278
- Knizek M. & Beaver, R. (2007) Taxonomy and systematics of bark and ambrosia beetles. En Lieutier et. al. (Ed.), *Insects in living trees in Europe, a Synthesis*. Springer The Netherlands
- Konstantinidis, K. T., & Rosselló-Móra, R. (2015). Classifying the uncultivated microbial majority: a place for metagenomic data in the Candidatus proposal. *Systematic and applied microbiology*, 38(4), 223-230.
- Kyre, B. R., Rodrigues, T. B., & Rieske, L. K. (2019). RNA interference and validation of reference genes for gene expression analyses using qPCR in southern pine beetle, *Dendroctonus frontalis*. *Scientific reports*, 9(1), 1-8.
- Lachowsky, L. E., Lalonde, R. G., & Reid, M. L. (2015). Can *Wolbachia* (Rickettsiaceae) explain female-biased sex ratios in mountain pine beetles (Coleoptera: Curculionidae)? *The Canadian Entomologist*, 147(6), 732-736.

- Lee, S., Kim, J., & Breuil, C. (2006). Diversity of fungi associated with mountain pine beetle, *Dendroctonus ponderosae*, and infested lodgepole pines in British Columbia. *Fungal Diversity*, 22, 91–105.
- Legendre, P., & Legendre, L. (2012). Cluster analysis. *Numerical Ecology*, 337–424. doi:10.1016/b978-0-444-53868-0.50008-3
- Lewinsohn, D., Lewinsohn, E., Bertagnolli, C. L., & Patridge, A. D. (1994). Blue-stain fungi and their transport structures on the Douglas-fir beetle. *Canadian Journal of Forest Research*, 24(11), 2275-2283.
- Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *bioinformatics*, 25(14), 1754-1760.
- Locey, K. J., & Lennon, J. T. (2016). Scaling laws predict global microbial diversity. *Proceedings of the National Academy of Sciences*, 113(21), 5970–5975. doi:10.1073/pnas.1521291113
- Loefer D. y Anderson, N. (2018) Impacts of the Mountain Pine Beetle on Sawmill Operations, Costs, and Product Values in Montana. *Forest Products J.* 68(1):15–24. doi:10.13073/FPJ-D-17-00041
- Lou, Q. Z., Lu, M., & Sun, J. H. (2014). Yeast diversity associated with invasive *Dendroctonus valens* killing *Pinus tabulaeformis* in China using culturing and molecular methods. *Microbial ecology*, 68(2), 397-415.
- Maguire, M., Kase, J. A., Roberson, D., Muruvanda, T., Brown, E. W., Allard, M., ... & González-Escalona, N. (2021). Precision long-read metagenomics sequencing for food safety by detection and assembly of Shiga toxin-producing *Escherichia coli* in irrigation water. *PloS one*, 16(1), e0245172.
- Massey, C.L., (1956). Nematode parasites and associates of the Engelmann spruce beetle (*Dendroctonus engelmanni* Hopk.). *P. Helm. Soc. Wash.* 23, 14–24.
- Massey, C. L. (1966). The nematode parasites and associates of *Dendroctonus adjunctus* (Coleoptera: Scolytidae) in New Mexico. *Annals of the Entomological Society of America*, 59(3), 424-440.
- Ma, B., Tromp, J y Li, M. (2002) "PatternHunter: faster and more sensitive homology search", *Bioinformatics*;18(3):440-5
- Ma, L. J., Geiser, D. M., Proctor, R. H., Rooney, A. P., O'Donnell, K., Trail, F., ... & Kazan, K. (2013). Fusarium pathogenomics. *Annual review of microbiology*, 67, 399-416.
- McCartney-Melstad, E., Gidiş, M., & Shaffer, H. B. (2019). An empirical pipeline for choosing the optimal clustering threshold in RADseq studies. *Molecular ecology resources*, 19(5), 1195-1204.
- McMurdie, P. J., & Holmes, S. (2013). phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PloS one*, 8(4), e61217.
- Mills, N. J. (1983). The natural enemies of scolytids infesting conifer bark in Europe in relation to the biological control of *Dendroctonus* spp. in Canada. *Biocontrol News and Information*, 4(4), 303-326.
- Mineeva, O., Rojas-Carulla, M., Ley, R. E., Schölkopf, B., & Youngblut, N. D. (2020). DeepMAseD: evaluating the quality of metagenomic assemblies. *Bioinformatics*, 36(10), 3011-3017.

- Morales-Jiménez, J., Zúñiga, G., Villa-Tanaca, L., & Hernández-Rodríguez, C. (2009). Bacterial community and nitrogen fixation in the red turpentine beetle, *Dendroctonus valens* LeConte (Coleoptera: Curculionidae: Scolytinae). *Microbial Ecology*, *58*(4), 879-891
- Morales-Jiménez, J., de León, A. V. P., García-Domínguez, A., Martínez-Romero, E., Zúñiga, G., & Hernández-Rodríguez, C. (2013). Nitrogen-fixing and uricolytic bacteria associated with the gut of *Dendroctonus rhizophagus* and *Dendroctonus valens* (Curculionidae: Scolytinae). *Microbial ecology*, *66*(1), 200-210.
- Nabors, M. (2006). Crecimiento vegetal secundario. En *Introducción a la Botánica*(119). Madrid, España: Pearson Educación, S.A..
- Nevill, R. J., & Safranyik, L. (1996). Interaction between the bluestain fungal associates of mountain pine, and pine engraver beetles, (*Dendroctonus ponderosae* and *Ips pini*, Coleoptera: Scolytidae) and their effects on the beetles. *Journal of the Entomological Society of British Columbia*, *93*, 39-48.
- Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'hara, R. B., ... & Oksanen, M. J. (2013). Package 'vegan'. *Community ecology package, version*, *2*(9), 1-295.
- Orantes, L. C., Monroy, C., Dorn, P. L., Stevens, L., Rizzo, D. M., Morrissey, L., ... Helms Cahan, S. (2018). Uncovering vector, parasite, blood meal and microbiome patterns from mixed-DNA specimens of the Chagas disease vector *Triatoma dimidiata*. *PLOS Neglected Tropical Diseases*, *12*(10), e0006730. doi:10.1371/journal.pntd.0006730
- Parks, D.H., Chuvochina, M., Waite, D. W., Rinke, C., Skarshewski, A., Chaumeil, P. A., & Hugenholtz, P. (2018). A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nature biotechnology*, *36*(10), 996-1004.
- Pearman, W. S., Freed, N. E., & Silander, O. K. (2020). Testing the advantages and disadvantages of short-and long-read eukaryotic metagenomics using simulated reads. *BMC bioinformatics*, *21*, 1-15.
- Pereira-Marques, J., Hout, A., Ferreira, R. M., Weber, M., Pinto-Ribeiro, I., van Doorn, L. J., ... & Figueiredo, C. (2019). Impact of host DNA and sequencing depth on the taxonomic resolution of whole metagenome sequencing for microbiome analysis. *Frontiers in microbiology*, *10*, 1277.
- Pérez Vera, O. A., Alvarado Rosales, D., Cibrián Tovar, D., Equihua Martínez, A., & Cárdenas Soriano, E. (2011). Hongos ophiostomatoides de galerías de *Dendroctonus adjunctus* Blandford en *Pinus hartwegii* Lindl. *Revista mexicana de ciencias forestales*, *2*(8), 35-46.
- Phillips, M. A., & Croteau, R. B. (1999). Resin-based defenses in conifers. *Trends in plant science*, *4*(5), 184-190.
- PO, M. P. B. I. (2006). Determining fungal diversity on *Dendroctonus ponderosae* and *Ips pini* affecting lodgepole pine using cultural and molecular methods.
- Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., ... & Glöckner, F. O. (2012). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic acids research*, *41*(D1), D590-D596.
- Quinlan AR and Hall IM, 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. *26* (6), pp. 841–842.

- Rosselló-Móra, R., & Amann, R. (2015). Past and future species definitions for Bacteria and Archaea. *Systematic and applied microbiology*, 38(4), 209-216.
- Reeve, J. D., Anderson, F. E., & Kelley, S. T. (2012). Ancestral state reconstruction for *Dendroctonus* bark beetles: evolution of a tree killer. *Environmental entomology*, 41(3), 723-730.
- Relman, D. A. (2008). 'Til death do us part': coming to terms with symbiotic relationships. *Nature Reviews Microbiology*, 6(10).
- Rinke, C., Chuvochina, M., Mussig, A. J., Chaumeil, P. A., Davín, A. A., Waite, D. W., ... & Hugenholtz, P. (2021). A standardized archaeal taxonomy for the Genome Taxonomy Database. *Nature Microbiology*, 1-14.
- Rivera, F. N., Gonzalez, E., Gomez, Z., Lopez, N., Hernandez-Rodriguez, C., Berkov, A., & Zuniga, G. (2009). Gut-associated yeast in bark beetles of the genus *Dendroctonus* Erichson (Coleoptera: Curculionidae: Scolytinae). *Biological Journal of the Linnean Society*, 98(2), 325-342.
- Roe, A. D., James, P. M. A., Rice, A. V., Cooke, J. E. K., & Sperling, F. A. H. (2011). Spatial Community Structure of Mountain Pine Beetle Fungal Symbionts Across a Latitudinal Gradient. *Microbial Ecology*, 62(2), 347-360.
- Rognes, T., Flouri, T., Nichols, B., Quince, C., & Mahé, F. (2016). VSEARCH: a versatile open source tool for metagenomics. *PeerJ*, 4, e2584.
- Rose, W. E. (1966). The biology and ecology of *Dendroctonus valens* Lec., and the biology, ecology, and control of *Dendroctonus frontalis* (= *mexicanus*) Zimm. in central Mexico (Coleoptera: Scolytidae).
- Samman, S y Logan J. (2000) Assessment and response to bark beetle outbreaks in the Rocky Mountain area. USDA Forest Service. *General Technical Report RMRS-GTR-62*
- Schweyen, H., Rozenberg, A., & Leese, F. (2014). Detection and Removal of PCR Duplicates in Population Genomic ddRAD Studies by Addition of a Degenerate Base Region (DBR) in Sequencing Adapters. *The Biological Bulletin*, 227(2), 146-160. doi:10.1086/bblv227n2p146
- Segata, N., Boernigen, D., Tickle, T. L., Morgan, X. C., Garrett, W. S., & Huttenhower, C. (2013). Computational meta'omics for microbial community studies. *Molecular systems biology*, 9(1), 666.
- Setiawan, R., Sumerta, I. N., Anita, S., & Narakusumo, R. P. (2017, September). Isolation and Enzymatic Activities of Bacteria Associated to Weevils (Coleoptera: Curculionidae). In *Proceedings of the 4th International Symposium on Innovative Bioproduction, Bogor, Indonesia* (pp. 39-47).
- Six, D.L. & Paine, T- (1996) *Leptographium pyrinum* is a mycangial fungus of *Dendroctonus adjunctus*, *Mycologia*, 88(5), 739- 744, DOI: 10.1080/00275514.1996.12026711
- Six, D. L., & Paine, T. D. (1998). Effects of mycangial fungi and host tree species on progeny survival and emergence of *Dendroctonus ponderosae* (Coleoptera: Scolytidae). *Environmental Entomology*, 27(6), 1393-1401.
- Six, D. L. (2003). Bark beetle-fungus symbioses. En Borutis, K. y Miller, T.A. (Eds.), *Insect symbiosis*, (99-116). Boca Raton: CRC Press
- Six, D. L., & Klepzig, K. D. (2004). *Dendroctonus* bark beetles as model systems for studies on symbiosis. *Symbiosis*.

- Six, D. L. (2012). Ecological and evolutionary determinants of bark beetle—fungus symbioses. *Insects*, 3(1), 339-366.
- Six, D. y Bracewellm R. (2015) *Dendroctonus*. En F.E.: Vega y R.W. Hofstetter (Eds.), *Barks Beetles, Biology and Ecology of Native and Invasive Species*.(305-350) Elsevier, EEUU
- Six, D. L. (2020). A major symbiont shift supports a major niche shift in a clade of tree-killing bark beetles. *Ecological Entomology*, 45(2), 190-201.
- Skelton, J., Jusino, M. A., Li, Y., Bateman, C., Thai, P. H., Wu, C., ... Hulcr, J. (2018). Detecting Symbiosis in Complex Communities: the Fungal Symbionts of Bark and Ambrosia Beetles Within Asian Pines. *Microbial Ecology*, 76(3), 839–850.
- Stamatakis, A. (2014) RAxML Version 8: A tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies. *Bioinformatics*. 30 (9), 1312-1313
- Sullivan, B. T., Grady, A. M., Hofstetter, R. W., Pureswaran, D. S., Brownie, C., Cluck, D., ... & Zúñiga, G. (2021). Evidence for semiochemical divergence between sibling bark beetle species: *Dendroctonus brevicomis* and *Dendroctonus barberi*. *Journal of chemical ecology*, 47(1), 10-27.
- Tanabe, A. S., & Toju, H. (2013). Two New Computational Methods for Universal DNA Barcoding: A Benchmark Using Barcode Sequences of Bacteria, Archaea, Animals, Fungi, and Land Plants. *PLoS ONE*, 8(10), e76910. doi:10.1371/journal.pone.0076910
- Tanyeli, E., Sevim, A., Demirbag, Z., Eroglu, M., & Demir, I. (2010). Isolation and virulence of entomopathogenic fungi against the great spruce bark beetle, *Dendroctonus micans*(Kugelann) (Coleoptera: Scolytidae). *Biocontrol Science and Technology*, 20(7), 695–701. doi:10.1080/09583151003717219
- Taylor, J. W. (2011). One fungus= one name: DNA and fungal nomenclature twenty years after PCR. *IMA fungus*, 2(2), 113-120.
- Thong, C. H., & Webster, J. M. (1975). Effects of *Contortylenchus reversus* (Nematoda: Sphaerulariidae) on hemolymph composition and oocyte development in the beetle *Dendroctonus pseudotsugae* (Coleoptera: Scolytidae). *Journal of invertebrate pathology*, 26(1), 91-98.
- Thong, C. H., & Webster, J. M. (1983). Nematode parasites and associates of *Dendroctonus* spp. and *Trypodendron lineatum* (Coleoptera: Scolytidae), with a description of *Bursaphelenchus varicauda* n. sp. *Journal of Nematology*, 15(2), 312.
- Twiddy, S. (2010) Life Cycle of Mountain Pine Beetle. Recuperado el 26/12/2020, de Sean Twiddy's Bughaus Productions Sitio web: [http://www.seantwiddy.com/?attachment\\_id=369](http://www.seantwiddy.com/?attachment_id=369)
- Valerio-Mendoza, O., García-Román, J., Becerril, M., Armendáriz-Toledano, F., Cuéllar-Rodríguez, G., Negrón, J. F., ... & Zúñiga, G. (2019). Cryptic species discrimination in western pine beetle, *Dendroctonus brevicomis* LeConte (Curculionidae: Scolytinae), based on morphological characters and geometric morphometrics. *Insects*, 10(11), 377.
- Vasanthakumar, A., Delalibera Jr, I., Handelsman, J., Klepzig, K. D., Schloss, P. D., & Raffa, K. F. (2006). Characterization of gut-associated bacteria in larvae and adults of the southern pine beetle, *Dendroctonus frontalis* Zimmermann. *Environmental Entomology*, 35(6), 1710-1717.

- Victor, J., & Zuniga, G. (2016). Phylogeny of *Dendroctonus* bark beetles (Coleoptera: Curculionidae: Scolytinae) inferred from morphological and molecular data. *Systematic Entomology*, 41(1), 162-177.
- Wang, X., Chen, H., Ma, C., & Li, Z. (2010). Chinese white pine beetle, *Dendroctonus armandi* (Coleoptera: Scolytinae), population density and dispersal estimated by mark-release-recapture in Qinling Mountains, Shaanxi, China. *Applied Entomology and Zoology*, 45(4), 557-567.
- Wang, J., Chen, H., & Tang, M. (2017). Community structure of gut bacteria of *Dendroctonus armandi* (Coleoptera: Curculionidae: Scolytinae) larvae during overwintering stage. *Scientific reports*, 7(1), 1-11.
- Wang, X., Chen, H., Ma, C., & Li, Z. (2010). Chinese white pine beetle, *Dendroctonus armandi* (Coleoptera: Scolytinae), population density and dispersal estimated by mark-release-recapture in Qinling Mountains, Shaanxi, China. *Applied Entomology and Zoology*, 45(4), 557-567.
- Warnes, M. G. R., Bolker, B., Bonebakker, L., Gentleman, R., & Huber, W. (2016). Package 'gplots'. *Various R programming tools for plotting data*.
- Werren, J. H. (1997). Biology of wolbachia. *Annual review of entomology*, 42(1), 587-609.
- White, R. A., Callister, S. J., Moore, R. J., Baker, E. S., & Jansson, J. K. (2016). The past, present and future of microbiome analyses. *Nature Protocols*, 11(11), 2049-2053.
- Wickham, H., Chang, W., & Wickham, M. H. (2016). Package 'ggplot2'. *Create elegant data visualisations using the grammar of graphics. Version*, 2(1), 1-189.
- Wommack, K. E., Bhavsar, J., & Ravel, J. (2008). Metagenomics: read length matters. *Applied and environmental microbiology*, 74(5), 1453-1463.
- Wood, S.L. (1982) The bark and ambrosia beetles of North and Central America (Coleoptera: Scolytidae), a taxonomic monograph. *Great Basin Naturalist Memoirs*, 6, 1–1359.
- Woos, S. (1982) A revision of the bark beetles genus *Dendroctonus* Erichson (Coleoptera:Scolytinae): a taxonomic monograph. *Great Basin Nat.* 6:1-1359
- Xu, L., Lou, Q., Cheng, C., Lu, M., & Sun, J. (2015). Gut-associated bacteria of *Dendroctonus valens* and their involvement in verbenone production. *Microbial ecology*, 70(4), 1012-1023.
- Xu, L. T., Lu, M., & Sun, J. H. (2016). Invasive bark beetle-associated microbes degrade a host defensive monoterpene. *Insect science*, 23(2), 183-190.
- Yaman, M., Radek, R., Weiser, J., & Aydin, Ç. (2010). A microsporidian pathogen of the predatory beetle *Rhizophagus grandis* (Coleoptera: Rhizophagidae). *Folia parasitologica*, 57(3), 233.
- Yaman, M., & Radek, R. (2015). *Mattesia weiseri* sp. nov., a new neogregarine (Apicomplexa: Lipotrophidae) pathogen of the great spruce bark beetle, *Dendroctonus micans* (Coleoptera: Curculionidae, Scolytinae). *Parasitology research*, 114(8), 2951-2958.
- Yang, Z. W., Men, Y., Zhang, J., Liu, Z. H., Luo, J. Y., Wang, Y. H., ... & Xie, Q. (2021). Evaluation of Sample Preservation Approaches for Better Insect Microbiome Research According to Next-Generation and Third-Generation Sequencing. *Microbial Ecology*, 1-10.

Yim, W. C., & Cushman, J. C. (2017). Divide and Conquer (DC) BLAST: fast and easy BLAST execution within HPC environments. *PeerJ*, 5, e3486.

Zúñiga, G., Cisneros, R., Hayes, J. L., & Macias-Samano, J. (2002). Karyology, geographic distribution, and origin of the genus *Dendroctonus* Erichson (Coleoptera: Scolytidae). *Annals of the Entomological Society of America*, 95(3), 267-275.

ZV. (2019) Hasta 60% del bosque dañó plaga del gorgojo. La Tribuna. Recuperado el 26 de abril de 2020 de <https://www.latribuna.hn/2019/12/12/hasta-60-del-bosque-dano-plaga-del-gorgojo/>

Anexo 5. Tabla de loci de todos los phyla con más de 90% de similitud recuperados por los diferentes ensambles

	Denovo _97	Denovo _95	DRef_ 97	DRef_ 95	DPond_ 97	DPond_ 95	GRef_ 97	GRef_ 95	GRef_ 97_m4	GRef_ 7_m2	GRef_ 90_m2	GRef_ 85_m2
Actinobacteria	45	2	30	26	1	1	0	0	0	0	0	0
Annelida	3	0	2	2	0	0	0	0	0	0	0	0
Apicomplexa	208	26	149	147	0	0	0	0	0	0	0	2
Aquificae	2	0	0	0	0	0	0	0	0	0	0	0
Arthropoda	167941	15618	10677	9433	248	232	31	29	36	41	24	22
Artverviricota	1	1	1	1	0	0	0	0	0	0	0	0
Ascomycota	170	21	91	92	1	1	3	3	4	292	350	370
Bacillariophyta	1	1	1	1	0	0	0	0	0	0	0	0
Bacteroidetes	107	16	84	77	0	0	0	0	0	0	0	0
Basidiomycota	28	5	13	12	0	0	0	0	0	0	0	0
Brachiopoda	3	0	3	2	0	0	0	0	0	0	0	0



Candidatus Diapherotrites	1	0	1	0	0	0	0	0	0	0	0	0
Candidatus Lokiarchaeota	1	0	1	1	0	0	0	0	0	0	0	0
Candidatus Woesebacteria	2	1	2	2	0	0	0	0	0	0	0	0
Cercozoa	1	1	0	0	0	0	0	0	0	0	0	0
Chlamydiae	1	0	1	0	0	0	0	0	0	0	0	0
Chlorobi	2	0	1	0	0	0	0	0	0	0	0	0
Chloroflexi	1	0	0	1	0	0	0	0	0	0	0	0
Chlorophyta	16	1	8	8	0	0	0	0	0	0	0	1
<b>Chordata</b>	4313	500	2429	2194	9	9	1	1	2	2	5	6
Chytridiomycota	2	0	0	1	0	0	0	0	0	0	0	0
Ciliophora	34	3	22	19	0	0	0	0	0	0	0	0
Cnidaria	45	1	26	26	0	0	0	0	0	0	0	0

Crenarchaeota	6	0	6	6	0	0	0	0	0	0	0	0
Cyanobacteria	33	5	19	16	2	0	0	0	0	0	0	0
Deinococcus-Thermus	1	0	0	0	0	0	0	0	0	0	0	0
Discosea	1	0	1	1	0	0	0	0	0	0	0	0
Duplornaviricota	1	0	0	0	0	0	0	0	0	0	0	0
Echinodermata	33	9	20	19	0	0	0	0	0	0	0	0
Elusimicrobia	1	0	1	1	0	0	0	0	0	0	0	0
Euglenozoa	15	1	7	7	0	0	0	0	0	0	0	0
Euryarchaeota	15	0	11	9	0	0	0	0	0	0	0	0
Evosea	35	3	23	17	0	0	0	0	0	0	0	0
<b>Firmicutes</b>	196	14	142	131	1	1	0	0	0	0	0	0
Fusobacteria	20	0	13	14	0	0	0	0	0	0	0	0
Hemichordata	3	0	1	0	0	0	0	0	0	0	0	0
Heterolobosea	1	0	1	1	0	0	0	0	0	0	0	0

Ignavibacteriae	1	0	0	1	0	0	0	0	0	0	0	0
Microsporidia	8	0	8	7	0	0	1	1	1	1	1	1
Mollusca	277	24	118	112	1	1	0	0	0	0	0	0
Mucoromycota	8	0	7	7	0	0	0	0	0	0	0	0
multiple	585	80	248	256	1	0	7	6	5	9	13	11
Negarnaviricota	2	1	2	2	0	0	0	0	0	0	0	0
Nematoda	463	59	257	227	8	7	0	0	0	0	0	0
Nemertea	2	0	2	1	0	0	0	0	0	0	0	0
no phylum	51	5	43	31	0	0	0	0	0	0	2	2
Nucleocytoviricota	14	4	7	6	0	0	0	0	0	0	0	0
Oomycota	17	0	3	2	0	0	0	0	0	0	0	0
Parabasalia	4	2	4	4	0	0	0	0	0	0	0	0
Peploviricota	1	0	1	1	0	0	0	0	0	0	0	0
Perkinsozoa	2	1	0	0	0	0	0	0	0	0	0	0

Pisuviricota	1	0	1	1	0	0	0	0	0	0	0	0
Planctomycetes	6	0	3	2	0	0	0	0	0	0	0	0
Platyhelminthes	134	9	69	58	1	1	0	0	0	0	0	0
Porifera	0	1	0	1	0	0	0	0	0	0	0	0
Preplasmiviricota	1	0	1	1	0	0	0	0	0	0	0	0
Priapulida	4	1	1	1	0	0	0	0	0	0	0	0
Proteobacteria	5543	155	4879	5031	2	2	0	0	0	1	1	2
Rhodophyta	10	1	5	6	0	0	0	0	0	0	0	0
Rotifera	1	0	1	0	0	0	0	0	0	0	0	0
Spirochaetes	19	2	10	9	0	0	0	0	0	0	0	0
Streptophyta	1998	236	1313	1189	4	4	1	0	1	1	0	0
Tenericutes	40	5	27	25	0	0	0	0	0	0	0	0
Thaumarchaeota	2	0	2	2	0	0	0	0	0	0	0	0
Thermodesulfobacteria	1	0	1	1	0	0	0	0	0	0	0	0

Thermotogae	6	1	1	1	0	0	0	0	0	0	0	0
Uroviricota	15	2	9	10	0	0	0	0	0	0	0	0
TOTAL	182505	16818	20810	19262	279	259	44	40	49	347	396	417